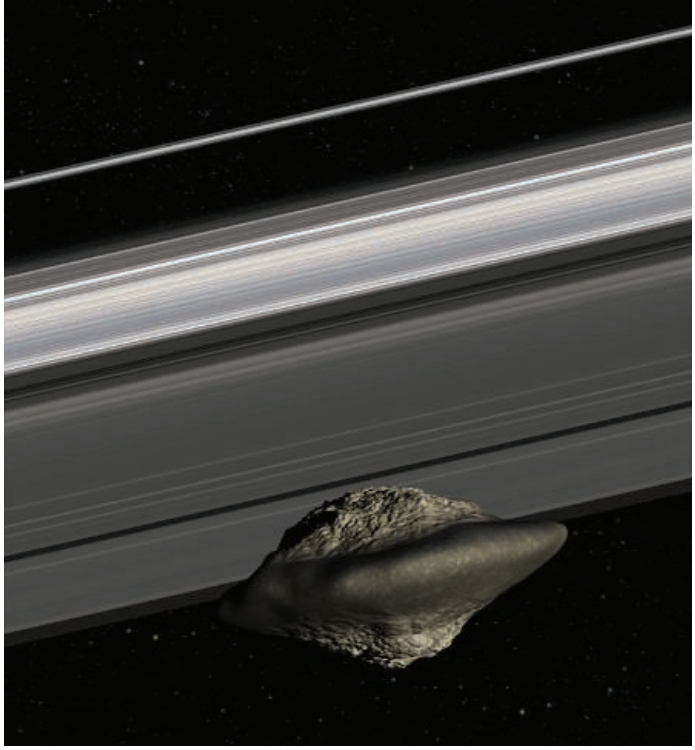
7 December 2007 | $10

Science

AAAS

EDITED BY STELLA HURTLEY AND PHIL SZUROMI



<< Satellites Take Shape

Saturn's smallest moons were thought to be the remnants of collisions between larger moons. These collisions would also have created debris that became further refined and ultimately filled in the rings that circle the planet. **Porco *et al.*** (p. 1602) instead show that the small moons grew out of accreted ring material. Their analysis of the shapes and densities of moons in images taken by the Cassini spacecraft indicates that the moons grew to a maximum size governed by the balance of local gravity. The sizes were also limited by availability of additional ring material once those regions were cleared out. Modeling by **Charnoz *et al.*** (p. 1622) reveal that this process results in the characteristic elongated and bulging shapes of two of the moons, Pan and Atlas.

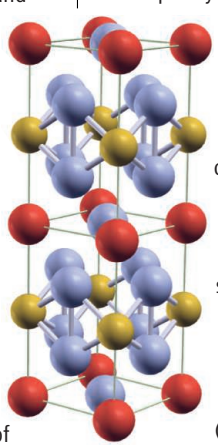
Optical Quantum Computing

There are currently several experimental routes being pursued in the goal of realizing a working computer. One approach is linear optics, in which the bits of quantum information (qubits) are stored in the polarization state of single photons and the logic gates are formed from simple elements such as beam splitters, mirrors, and waveplates. **O'Brien** (p. 1567) reviews recent progress in this area and points out the challenges that remain to be addressed.

Evidence of Cosmic Texture

After the Big Bang, the expanding universe progressed through a series of phase changes in which various forces and fields became decoupled and symmetries were broken. According to theory, these phase changes spread at the speed of light across patches of the universe. However, they may have done so irregularly and left behind cosmic defects similar to those seen in crystals, although none has yet been seen by astronomers. **Cruz *et al.*** (p. 1612, published online 25 October; see the Perspective by **Brandenberger**) propose that the remnant of a cosmic defect known as a texture has the right properties to explain an unusual cold spot in the cosmic microwave background, the frozen map of the universe at the point when photons and the first atoms decoupled hundreds of thousands of years after the Big Bang. If this feature is a texture, then its existence constrains the fundamental symmetry-breaking energy scale.

CREDITS (TOP TO BOTTOM): FRÉDÉRIC DURILLON/WWW.ANIMEA.COM; SHIM *ET AL.*



Modeling Electron Correlation

Capturing the essential physics underlying highly correlated electron systems is a huge challenge in contemporary condensed matter physics. However, the shear complexity in describing all the interactions that can take part has meant a full understanding has been lacking. **Shim *et al.*** (p. 1615, published online 1 November; see the Perspective by **Fisk**) introduce a theoretical effort to understand the angle-resolved photoemission spectra and optical spectra of the prototypical heavy fermion compound $CeIrIn_5$ in terms of dynamical mean-field theory combined with local density approximation (DMFT+LDA) calculations. Their calculations show how the electrons evolve with decreasing temperature from a localized high-temperature state to a delocalized fluid of quasiparticles that have masses many times greater than that of a free electron.

Making the Right Cut

Regulated intramembrane proteolysis (RIP) represents an important signaling mechanism that is conserved from bacteria to humans. A notable example of RIP is the activation by cleavage of the transcription factor Sterol Regulatory Element Binding Proteins by site-2 protease (S2P), a key event in regulating cellular cholesterol levels. **Feng *et al.*** (p. 1608) now present the crystal structure of the transmembrane core of an archaebacterial S2P metalloprotease, which provides insight into how S2P functions. The structure shows the mechanism of cleavage at an active site, containing a catalytic zinc ion that is embedded deep in the membrane. Two conformations observed in the crystals suggest that a helical gating mechanism controls substrate access.

Oil-Repelling Surfaces

Several approaches have been used to make superhydrophobic materials that excel at repelling water. These methods typically combine a material with a low surface energy with a form of surface roughness that keeps the water as buoyant droplets on the surface with high contact angles. It is much more difficult to use this route to create superoleophobic materials because organic liquids typically have much lower surface tensions. Indeed, calculations indicate that it might not be possible to achieve such a surface from just these two design criteria. **Tuteja *et al.*** (p. 1618) show that by considering a third factor, the use of reentrant surface structures (which include concave surface features), surfaces can be created that repel a wide range of organic materials, including octane and decane.

Avoiding Resistance

One of the most commonly used classes of pesticides are toxins produced by the bacteria *Bacillus thuringiensis*, known as Bt toxins. The

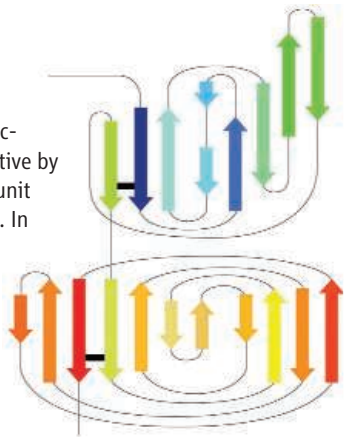
Continued on page 1521

Published by AAAS

Continued from page 1519

primary threat to the efficacy of Bt toxins is the evolution of resistance by pests. In major pests, resistance to Bt toxins in the Cry1A family is linked to alterations in cadherin proteins that act as primary toxin receptors in the midgut of susceptible insects. Now **Soberón *et al.*** (p. 1640, published online 15 November; see the Perspective by **Moar and Anilkumar**) show that engineered Bt toxins can overcome insect resistance. Furthermore, in investigating the effectiveness of these engineered proteins, the authors provided evidence for the means by which these toxins cause mortality in insects.

Bacterial Pilus Structure Revealed



Bacterial pili, filamentous adhesive structures that extend from the cell surface, are important virulence factors and potential vaccine targets. Pili from Gram-negative bacteria have been structurally characterized. Now **Kang *et al.*** (p. 1625; see the Perspective by **Yeates and Clubb**) describe the structure of the major pilin subunit from a Gram-positive human pathogen, *Streptococcus pyogenes*. In the crystal, the subunits associate in columns reminiscent of the likely arrangement in native pili. The structure also reveals intramolecular isopeptide bonds that may stabilize the structure and contribute to protease resistance. This could be a more general mechanism of protein stabilization in Gram-positive organisms, which lack the disulfide bond formation machinery of Gram-negative bacteria.

Dissecting X Inactivation

One of the two X chromosomes in mammalian females is randomly inactivated early in development to match the single active X chromosome of males. This process is regulated through the X-inactivation center (Xic). The two Xics interact in *trans* at the beginning of X-inactivation, presumably to allow reciprocal activation/inactivation. So far, single copies of elements from the Xic have not been able to recapitulate X inactivation, suggesting additional elements must be required. **Augui *et al.*** (p. 1632) find that a region ~200 kilobases upstream of the Xic— the X-pairing-region (*Xpr*) —is sufficient in a single copy to allow a transient interaction between the two Xics at a time before the beginning of X inactivation. This pairing is cell cycle dependent, can occur from an ectopic location, and may activate the expression of *Xist*, a noncoding RNA that coats the inactive X chromosome.

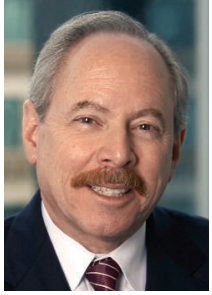
Dealing with DNA Damage

Like railway tracks severed clean through, a break in both strands of genomic DNA can result in disaster. To avoid potential cellular chaos in the face of such damage, a complex DNA damage response has evolved. In eukaryotes, phosphorylation of histone H2AX and polyubiquitination of proteins at sites of damage recruit DNA-repair proteins, forming cytologically visible foci. **Kolas *et al.*** (p. 1637, published online 15 November) show that the ubiquitin ligase RNF8 is responsible for polyubiquitination at double-strand breaks in yeast. RNF8 is recruited to the damaged sites though its interactions with phosphorylated MDC1, and acts downstream of MDC1 to promote the formation of at least two distinct classes of repair foci. RNF8 binds the E2 conjugating enzyme UBC13 to drive polyubiquitination at the site of the break, and also helps regulate the G2/M cell-cycle checkpoint.

Learn from Your Mistakes

Human experience is based on learning that our actions affect subsequent positive or negative outcomes. Rewards strengthen associations between contextual stimuli and actions thereby reinforcing and maintaining successful behavior; whereas punishments induce avoidance of erroneous actions. While we usually learn from both positive and negative reinforcement, the relative amount of learning from success or errors varies between individuals. **Klein *et al.*** (p. 1642) investigated a human genetic polymorphism associated with the density of brain dopamine D2 receptor. Reduced D2 receptor density was associated with less efficient learning from errors. In people with lower D2 receptor density, the reduced capacity to learn from errors was accompanied by reduced feedback-related activity in the posterior medial frontal cortex, an area known to monitor for negative action outcomes.

CREDIT: KANG *ET AL.*



Alan I. Leshner is Chief Executive Officer of AAAS and Executive Publisher of *Science*.

Vaughan Turekian is Chief International Officer of AAAS.

Chinese Science on the Move

ONE OF THE WORLD'S FASTEST-GROWING ECONOMIES IS ALSO THE WORLD'S fastest-growing supporter of scientific R&D. As of last year, China had moved to third place in the world in national R&D funding, closing fast on Japan, which was number two behind the United States. Not surprisingly, this rapid development has been garnering worldwide attention but also some concern. Although there are people who feel threatened by any competition, others are more reasonably worried that China's continuing rise may be at the expense of broader global societal issues, such as environmental protection, natural resource use, and research integrity. But overall, the rapid rise in Chinese science is good for China and the rest of the world, and the global scientific enterprise should do all it can to help.

As with all developing countries, recent progress in Chinese science has not always been smooth. Entire systems for local science and international cooperation had to be developed and are still evolving. But the rapid increase in R&D investment—with an annual growth rate of 18% over the past 5 years (the United States, Japan, and the European Union grew at a combined average rate of about 2.9%)—reflects a clear understanding by China's top political leadership that science and technology (S&T) are critical to their nation's future. This is not surprising because so many Chinese leaders are scientists and engineers by training. Educated as an engineer, Chinese leader Hu Jintao emphasizes the importance of investing in S&T in virtually every policy address. He included in his 2006 list of "do's and don'ts" for the Chinese populace: "Uphold science; don't be ignorant and unenlightened."

It's not just about increasing investment. China has long encouraged young researchers to go abroad for training, and many have stayed in their adopted countries. The Chinese government is now working hard to recruit the brightest and best-educated back to their homeland with job opportunities and state-of-the-art facilities and equipment. On a recent visit to Zhejiang University, we saw that laboratory equipment in virtually every discipline was equal in quality, if not yet in abundance, to that anywhere in the world.

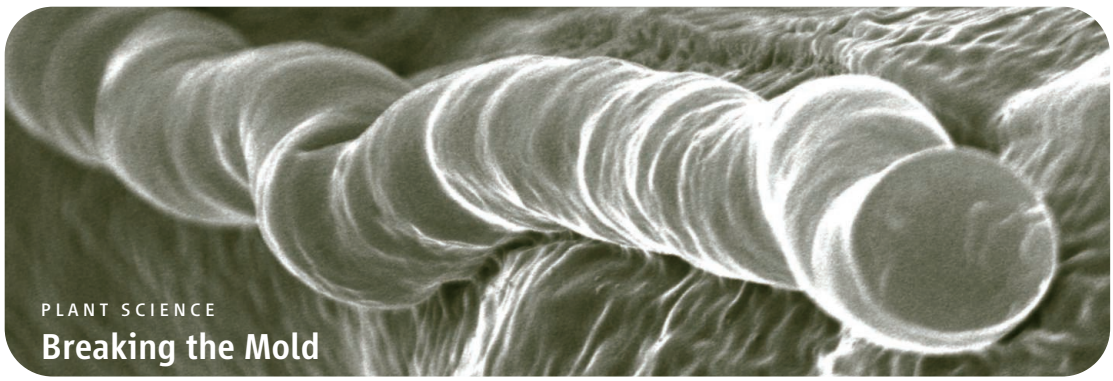
China knows that integration into the world scientific community is key to its future success as a nation, and Chinese science leaders clearly welcome partnerships to ease this process. At the end of September, the American Association for the Advancement of Science (AAAS) and the China Association for Science and Technology held a conference in Beijing on "Scientists' Social and Ethical Responsibilities" to explore how ethics standards and regulations could be aligned across countries and regions of the world. Minister of Science and Technology Wan Gang recently called on Chinese universities, government, and scientific journals to join an Advisory Committee on Research Integrity. These actions should inspire further initiatives that will ease formal collaborations between Chinese scientists and those in other countries.

This progress is good for China and for the rest of the world. S&T are embedded in every major societal problem, from the spread of infectious diseases, to environmental sustainability, to alleviating poverty. Many countries, including the United States, Canada, Australia, members of the European Union, and some in Africa, already have scientific partnerships with China. These collaborations are likely to grow, because they are focused in disciplines that have potential influence not only in each partner's country but internationally, including nanotechnology, biotechnology, and pharmaceutical development.

As China and other rapidly developing countries increase investment in S&T, researchers from other countries should help smooth the emergence of local science communities and their integration into the global science enterprise by reaching out for partnerships and collaboration. S&T are the foundation for innovation, economic growth, and quality of life in all parts of the world. And although there will be pressure in some circles to focus inward, it would be "unenlightened" and, in fact, counterproductive to view integration across the global science community as anything but helpful to all.

–Alan I. Leshner and Vaughan Turekian

10.1126/science.1153120



PLANT SCIENCE

Breaking the Mold

Filamentous fungi are believed to grow exclusively by means of elongation at their tips. However, this mechanism has seemed incompatible with the transmission of *Neoptyphodium* and *Epichloë* endophytic fungi through the embryos of seeds of their host grasses. By examining the vegetative growth of fungi within *Lolium* spp. grasses, Christensen *et al*. find evidence for hyphal expansion by physical expansion (intercalary growth); the fungi attach to their host cells (see hypha attached to a fescue leaf, above) and stretch within the space between cells, paralleling the leaf expansion growth mode of their host. Furthermore, the authors suggest that intercalary growth is an adaptation that maintains the integrity of hyphae within the leaves of its host. — LMZ

Fungal Genet. Biol. 10.1016/j.fgb.2007.07.013 (2007).

PSYCHOLOGY

Showing the Flag

Linking positions or policies to a country's flag by means of appeals to patriotism has become a potent weapon in the arsenal of political operatives of all stripes. Nuanced and multifaceted discussion can then be replaced by simplified yes/no choices, thus eliminating any middle ground and polarizing the voting public. Nevertheless, Hassin *et al.* show that in some situations, recourse to national symbols, such as the flag, can elicit the prosocial effect of drawing the citizenry from the extremes into the center. They found that presenting the Israeli flag subliminally—that is, too briefly for participants to become consciously aware that they had seen it—induced both right- and left-wing Israelis to adopt more moderate positions with respect to various aspects of the Israeli-Palestinian conflict. In addition, the consequences of this brief, undetected glimpse of the flag were reflected not only in the participants' declared voting intentions (in the elections of March 2006), but also in their actual voting behaviors. — GJC

Proc. Natl. Acad. Sci. U.S.A. **104**, 10.1073/pnas.0704679104 (2007).

CLIMATE SCIENCE

Cooler in the Forest

Human activity affects climate in many, sometimes opposing, ways. For example, emissions leading to increased atmospheric greenhouse gas concentrations cause warming, whereas those that raise the aerosol burden cause cooling. Moreover, our overall impact depends not only on what we send directly into the atmosphere but also on how we modify the land. Many analyses have concentrated solely on how land use change alters Earth's albedo, a measure of sunlight reflectivity at the surface, but changes in vegetation induce corresponding variations in moisture and heat fluxes that can have large impacts as well. In an effort to determine the effects on temperature of ecophysiological changes due to land use change in the southeastern United States, Juang *et al.* analyzed heat and radiation flux data collected via meteorological towers located in three distinct ecosystems: a grass-covered old field, a planted pine forest, and a hardwood forest. They found

that although the effect of albedo differences among the different ecosystems was large, with warming of nearly 1°C for the transition from old field to forest areas, the ecophysiological and aerodynamic effects of the same transitions could produce even greater cooling, of >2°C. Thus, contrary to some assertions, conversion of open fields to wooded fields will not necessarily make the world a hotter place. — HJS

Geophys. Res. Lett. **34**, L21408 (2007).

ECOLOGY/EVOLUTION

A Loss of Bivalves

The potential for ecosystems to shift abruptly from one state to another is becoming increasingly recognized, especially in aquatic environments. Cloern *et al.* document an unusual and instructive example in one ecosystem, brought about by changes in another, neighboring system. In 1999, the San Francisco Bay—a large lagoon-like estuary on the west coast of the United States—began to experience massive algal blooms for the first time since monitoring had begun more than two decades earlier. Such blooms are normally associated with eutrophication: the enrichment of waters by runoff of excessive nutrients (especially N and P) from agricultural land. However, in this case the nutrient loading of the estuary had been decreasing before the bloom. It appears that the bloom was instead the result of a collapse in the population of the bivalve consumers of the algae. This collapse was brought about by an influx of flatfish and crustacean predators of the bivalves into the estuary from the coastal ocean, which itself had resulted from a physical oceanographic change in the California Current System. Increased coastal upwelling of cold, nutrient-laden waters led to increased oceanic primary production and a bonanza for consumers and their predators, which

Continued on page 1527

CREDITS (TOP TO BOTTOM): CHRISTENSEN *ET AL.*, *FUNGAL GENET. BIOL.* 10.1016/J.FGB.2007.07.013 (2007); GABRIEL KATUL, *GEOPHYS. RES. LETT.* **34**, L21408 (2007)

Published by AAAS

Continued from page 1525

were recruited in substantial numbers into the neighboring estuary. Hence, the state change in the estuary was caused indirectly by hitherto-unsuspected connectivity with the ocean. — AMS

Proc. Natl. Acad. Sci. U.S.A. **104**, 18561 (2007).

APPLIED PHYSICS

Pick a Color

The ability to detect single photons makes it possible to investigate the quantum properties of light and to implement strategies for quantum cryptography and quantum communications with single photons as the information carriers. To date, photon detectors have come in two guises: They can be designed either for sensitivity at a single energy or over a broad range of energies, but neither option has offered on-chip tunability of the detected wavelength. Gustavsson *et al.* now describe a frequency-tunable single-photon detector for the microwave regime using a double quantum dot structure. They are able to shift the discrete energy levels of one dot with respect to the other by application of appropriate gate voltages. Using time-resolved charge detection techniques, they can then directly relate the detection of a tunneling electron to the absorption of a single photon, the energy of which corresponds to the tuned energy-level separation between the two dots. — ISO

Tunable single photon detector.

CREDIT: GUSTAVSSON *ET AL.*, *PHYS. REV. LETT.* **99**, 206804 (2007)

Phys. Rev. Lett. **99**, 206804 (2007).

CHEMISTRY

Gains in Contact

Self-assembled monolayers (SAMs) of alkanethiolates on gold or silver substrates have broad applications in surface patterning, but their flexible and thermally sensitive nature complicates efforts to probe properties such as conductivity. Mercury has been applied to SAM surfaces to form contact electrodes, but beyond its toxicity, its tendency to spread through flow can lead to short-circuiting and lack of measurement precision. Chiechi *et al.* show that a fluid eutectic composed of three parts gallium and one part indium by weight (dubbed "EGaIn") is a practical alternative. The primary advantage of this material is its capacity to retain its shape below the comparatively high applied surface stress of 0.5 N/m. The authors extruded droplets from a syringe onto a silver surface and then drew back the needle until a sharp micron-scale conical tip formed, the size of which could be tuned by varying the pulling rate. They could then apply these conducting tips to SAM surfaces for robust measurements of current densities as a function of applied voltage. Additional advantages of EGaIn include its stability in air and low toxicity. — JSY

Angew. Chem. Int. Ed. **46**, 10.1002/anie.200703642 (2007).

Science stke AAAS

www.stke.org

<< Food for Thought

Parkinson's disease is characterized by the loss of dopaminergic neurons in the substantia nigra pars compacta (SNc) of the brain. The reduced production and function of neurotrophins (proteins that promote neuronal survival) in patients' brains as compared to those of unaffected individuals may be a contributing factor in Parkinson's disease. Leptin is a hormone that functions in the hypothalamus to reduce appetite. Leptin receptors are also abundant in dopaminergic neurons in the SNc, leading Weng *et al*. to investigate whether leptin might play a role in neuronal survival. Immunohistochemical analyses showed that degeneration of mouse dopaminergic neurons in the SNc caused by injection into the brain of the neurotoxin 6-OHDA (a mouse model of Parkinson's disease) was less severe if the mice were pre-injected with leptin. Leptin was also protective against 6-OHDA–induced toxicity in mouse MN9D cells, a dopaminergic cell line. Western blotting assays and treatments with pharmacological inhibitors and short hairpin RNAs showed that the ability of leptin to block 6-OHDA–induced apoptosis was dependent on leptin receptor–mediated activation of Janus kinase 2, mitogen-activated or extracellular signal-regulated protein kinase kinase, extracellular signal–regulated kinase 1 and 2, and the transcription factor cAMP-response element binding protein (CREB), a known neuronal survival factor. Leptin also stimulated the phosphorylation and nuclear localization of CREB in dopaminergic SNc neurons and increased the abundance of brain-derived neurotrophic factor (BDNF) in the brain as compared to that in untreated animals. Together, these data suggest that treatment with leptin may be useful in therapies to combat Parkinson's disease. — JFF

J. Biol. Chem. **282**, 34479 (2007).

Published by AAAS

Olympic Archaeology

For more than 3 years, archaeologists in Beijing have been following the bulldozers as they tear into the land to make way for Olympic venues.

Last year, the bulldozers uncovered a particularly striking find: an ancient burial ground dating as far back as 1000 years that holds the

largest group of eunuch tombs ever unearthed in China. The 163 tombs of late Ming Dynasty (circa 1600) eunuchs came to light as workers were renovating the Clay Target Field of the Olympic Shooting Range.

Historian Leung Siu-kit of the University of Hong Kong calls the finds "exciting." Some Ming eunuchs were wealthy and high-ranking, and they carried great influence in the political intrigues of the day. Leung notes that "the epitaphs discovered in the tombs should provide more comprehensive information on eunuchs' careers and participation in the government" than has recorded history, which tends to focus on their misdeeds. Plans are to reconstruct the most elaborate tomb (see photo, left) at the Beijing Eunuch Culture Museum and open it in time for Olympic visitors.

Divorce Is Like an SUV

Divorce can be a headache—for Earth and not just the unhappy couple. The global surge in divorce rates adds up to more small and thus environmentally inefficient households, according to a study published online this week in the *Proceedings of the National Academy of Sciences*.

The proportion of U.S. families headed by divorced persons jumped from 5% to 15% between 1970 and 2000. According to the study, such households spent 46% more on electricity and 56% more on water per person. In addition, data from 12 countries indicate that divorcé(e)-headed households hold about one-third fewer people but occupy 33% to 95% more rooms per capita—38 million more rooms just in the United States.

"A lot of people … assume that human impact on the environment will … decline" with population declines, says co-author Jianguo Liu, an ecologist at Michigan State University in East Lansing. But "[with] decreasing household size, environmental impact may continue to increase." Sustainability researcher Manfred Lenzen of the University of Sydney, Australia, agrees that divorce can dramatically increase resource use but notes that some of that increase might result from higher divorce rates among the rich.

When Worlds Collide >>

Our solar system may have plenty of cosmic cousins. Scientists studying archived data have spotted an adolescent sunlike star with a dusty belt that shows evidence of the creation and violent destruction of baby planets. "There is no doubt that they are detecting the dusty debris of rocky [Earth-like] planet formation," says Scott Kenyon of the Harvard-Smithsonian Center for Astrophysics. A report of the find, by a team headed by Joseph Rhee of the University of California, Los Angeles, is in press at *The Astrophysical Journal*.

Artist's concept of how Earth-sized planets might collide in their formative years around a star.

Until 2005, astronomers had observed only very young possible planet-forming systems. Then data from the retired Infrared Astronomy Satellite revealed a more mature system, bolstering predictions that collisions continue well after planets form. The latest observation, from a star called HD 23514 in the Pleiades cluster, should "help generalize the model of planetary formation," says David Trilling of the University of Arizona in Tucson. Combined, the two discoveries allowed the team to estimate that about 1 in 1000 stellar systems share our system's turbulent past—and could share its present architecture.

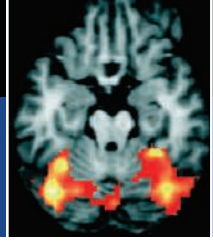
VIVE LE HOBBIT

A life-sized reconstruction of *Homo floresiensis*, a.k.a. the hobbit, will go on display on 11 December at the Musée de l'Homme in Paris. The model of the 18,000-year-old female, whose bones were discovered in 2003 on the Indonesian island of Flores, was created by French anthropological sculptor Elisabeth Daynès, aided by three anthropologists. The team relied on a three-dimensional stereolithograph of the hobbit's skull as well as publications on the skeleton.

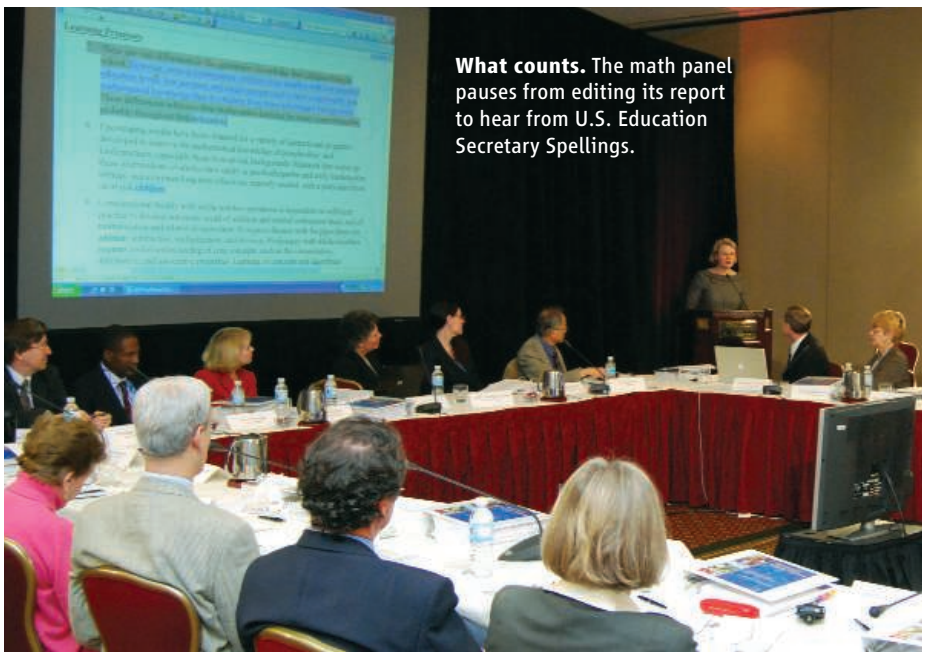
Anthropologist William Jungers of Stony Brook University School of Medicine in New York state says that the model is "very impressive and visually arresting." Nevertheless, Jungers and Stony Brook anatomist Susan Larson say the reconstruction fails to capture new data on the hobbit's peculiar shoulder anatomy (*Science*, 19 May 2006, p. 983). Larson says her work on the hobbit suggests that the shoulder blades were "positioned more on the sides of the rib cage than on the back," as they are in modern humans. Thus, the model's shoulders should have been more hunched, *H. erectus*–like. Daynès says that she was not aware of this work when she sculpted the hobbit in July but will include the data in any future reconstructions.

CREDITS (TOP TO BOTTOM): 2007 PHILIPPE PLAILLY/EURELIOS; BEIJING MUNICIPAL ADMINISTRATION OF CULTURAL HERITAGE; GEMINI OBSERVATORY/LYNETTE COOK

Thailand restricts some research
1537



Learning from mistakes
1539



What counts. The math panel pauses from editing its report to hear from U.S. Education Secretary Spellings.

CREDIT: U.S. DEPARTMENT OF EDUCATION

MATHEMATICS EDUCATION

U.S. Expert Panel Sees Algebra As Key to Improvements in Math

BALTIMORE, MARYLAND—No single report will end the decade-long debate about why U.S. students aren't doing better in math. But last week, a panel of experts assembled by the Department of Education signaled it had reached consensus on one of the most important topics in that debate: how students can become proficient in algebra.

Usually offered in the 8th or 9th grade, algebra is a gateway course for high school mathematics; without mastering algebra, a college degree in science or engineering is impossible. Its importance has made it the primary focus of the National Mathematics Advisory Panel, convened in April 2006. Last week, the group of 19 mathematicians, psychologists, and educators vetted a 68-page draft report due out this winter that members hope will play a major role in shaping math instruction across an education system that comes in 50 state flavors, with variations by 14,000 local school districts.

The report, debated line by line during an open 6-hour meeting at an airport hotel here, contains dozens of recommendations on how to boost student achievement in math. Taking aim at watered-down courses, the report defines the content of a rigorous algebra course as well as what students need to know before taking it. It urges school districts "to avoid an approach that continually revisits topics, year after year, without closure," part of what critics deride as a "mile-wide, inch-deep" math curriculum. It recommends giving teachers more authority to choose those educational materials and practices best suited to their students. It also calls for more useful assessments of what students know and for shifting educational policy debates "away from polarizing controversies."

At the same time, says panel chair Larry Faulkner, a chemist and former University of Texas president, the report will note that little or no good data exist on several hot-button issues. On choosing between a prescribed math curriculum presented by the teacher and one that incorporates what piques the interest of students, Faulkner notes, "it's a matter of religion, and it's important for the world to know that." That uncertainty is also true, he says, for whether elementary school students should be taught by math specialists rather than their regular classroom teacher. On the use of calculators in class, the group was deliberately equivocal: Math educator Douglas Clements of the University at Buffalo, State University of New York, told his fellow panelists that "we found limited to no impact on computational skills, problem-solving abilities, and conceptual development."

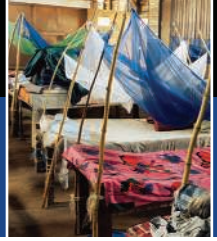
Despite the panel's desire for a consensus document, many issues seem likely to remain contentious long after the report is released. Take the discussion about how to teach arithmetic and whole numbers. Harvard University mathematician Wilfried Schmid argued strongly for including the phrase "the standard" in a paragraph that calls for "fluency with the standard algorithms for addition, subtraction, multiplication, and division." The two words, especially the article, are a rallying cry for the back-to-basics movement, which cites changes in the mathematics curriculum introduced in the 1990s as a major reason for low test scores. "Without that word," Schmid exhorted his colleagues, "we are sending a message that anything goes."

Math educator Deborah Ball of the University of Michigan, Ann Arbor, demurred, arguing that retaining the phrase would hamstring teachers who may want to use student-derived approaches in their lessons. "We're not talking about how to teach math in this paragraph," she explained, "and the use of alternative algorithms can be a useful tool for teachers. I'd like to drop the 'the.' " After more discussion, her view was adopted unanimously.

The vote was a cue for Francis "Skip" Fennel, president of the National Council of Teachers of Mathematics and chair of the subgroup that had worked on this section and who supported Ball's position, to take a coffee break. But the discussion wasn't over. As a way to reopen the issue, Schmid said another panel member, Fairfax, Virginia, middle school math teacher Vern Williams, had asked for his reaction to the vote and that "I am not distraught, but I'd be happier if the word were kept." The panel immediately took a second vote and decided, by a margin of 8–3, with three abstentions, to retain the article. Fennel then walked back in the room and, upon hearing about the new tally, declared: "You mean I lost?"

In addition to embodying the tensions ▶

Clues to a mysterious culture
1540



Malaria and the E word
1544

within the math community, the panel is also carrying some heavy political baggage. U.S. Education Secretary Margaret Spellings dropped by the meeting to give the panel a brief pep talk and urge it to finish quickly. Notwithstanding the panel's remaining work—it got through barely half of the 45 paragraphs in its draft executive summary—Spellings was comfortable describing its take-home message later to a small group of reporters.

The report will tell the country "what works" in math education, Spellings explained. "Once we know what works, it's our responsibility to align the resources" from the federal, state, and local governments. Spellings said the report's most important points are the need for students to master fractions, the importance of early childhood education, and the value of developing teacher skills, both during their training and after they are hired. Those messages dovetail with several initiatives proposed by the Bush Administration, including a $250 million Math Now program for middle school students that Congress has so far refused to fund.

–JEFFREY MERVIS

CLINICAL TRIALS

Gene Transfer an Unlikely Contributor to Patient's Death

Winding up an investigation into the mysterious death in July of a 36-year-old woman in a gene therapy safety trial, an expert panel this week concluded that the gene transfer was unlikely to have contributed to the tragedy but that this "cannot definitively be ruled out." Despite "an extraordinary effort," said chair Howard Federoff of Georgetown University in Washington, D.C., "we still are missing key pieces of information" needed to answer the question asked by the patient's widower: Would she be alive today if she had not taken part in the trial?

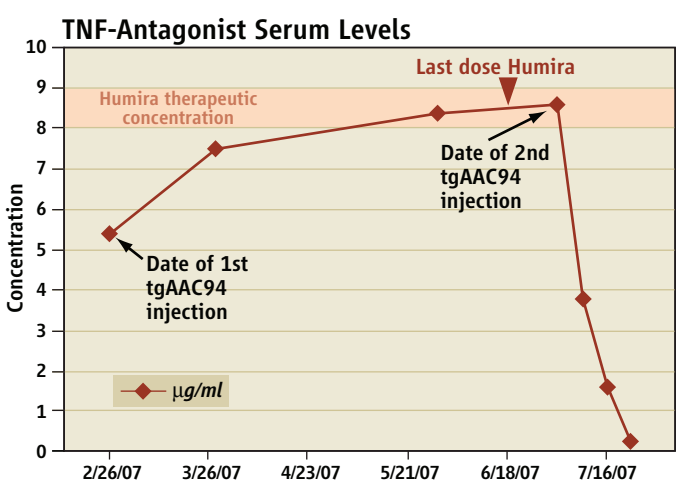
Just a week before this meeting of the Recombinant DNA Advisory Committee (RAC), the trial's sponsor, Targeted Genetics Corp. in Seattle, Washington, announced that its gene therapy treatment "did not contribute to the patient's death" and noted that the Food and Drug Administration (FDA) had lifted its hold on the trial (*Science*, 30 November, p. 1363). But issues raised by RAC may linger over the field of gene therapy, which had been blamed for two deaths since 1999.

The controversy concerns Jolee Mohr of Taylorville, Illinois, who died on 24 July, 3 weeks after receiving a second experimental gene therapy injection for rheumatoid arthritis in her knee. In September, RAC noted that Mohr apparently died from a fungal infection called histoplasmosis and a large blood clot. Mohr was taking an arthritis drug, Humira, which blocks a proinflammatory cytokine called tumor necrosis factor α (TNF-α) and suppresses the immune system. The protein produced by the gene therapy is also a TNF-α blocker, and if it spread beyond Mohr's knee, the combination with Humira may have left her vulnerable to the fungus.

Additional data "do not support [that] theory," RAC concluded this week. The level of TNF-α blocker detected in Mohr's blood was within the range expected from the dose of Humira she was taking, and it dropped after she received the gene therapy injection on 2 July (see graph). However, RAC recommended that Targeted Genetics develop an assay to distinguish between systemic TNF-α blockers like Humira and the gene therapy product.

RAC members also examined the possibility that Mohr had an immune reaction to the gene therapy vector, adeno-associated virus (AAV). Vector DNA appeared in other tissues only at extremely low levels, suggesting that the AAV did not replicate. However, RAC pointed to another possibility: that Mohr's immune system reacted to the virus's protein shell. That could have been tested by measuring certain T cells in Mohr's blood, but no whole blood samples were available. As a result, "an immune response cannot be *definitely* ruled out," RAC said. It recommended that all ongoing AAV trials monitor T-cell responses.

Even if the vector did play a role, "it was very unlikely to have been a significant contributor" to Mohr's illness and "was not the cause of her death," which was "primarily" from histoplasmosis with Humira as a risk factor, RAC concluded. FDA official Daniel Takefman said at the RAC meeting that the agency is "in agreement" with RAC's conclusion.



CREDIT: ADAPTED FROM OFFICE OF BIOTECHNOLOGY ACTIVITIES, NIH

No smoking gun. Patient Jolee Mohr's response after receiving the drug Humira and the gene therapy product tgAAC94 did not exceed an expected "therapeutic concentration."

Defending the company's announcement that its therapy was not to blame, Targeted Genetics Executive Vice President and Chief Scientific Officer Barrie Carter points out that the data and safety monitoring board for the trial, an independent group, concluded that the death was not related to the trial. That is not inconsistent with RAC's findings, Carter says; the problem is, "you can't prove a negative."

The company now plans to resume the safety trial of 127 patients, but, in keeping with RAC and FDA recommendations, it will not give a second dose to patients if they have a fever, as Mohr did, or show other signs of infection.

–JOCELYN KAISER

ACADEMIC FREEDOM

Thai Science Agency Clamps Down on Sensitive Research

Concerned about Thailand's image and security, authorities are preparing to restrict foreign research involving three touchy subjects: child labor, prostitution, and a simmering Islamic insurgency. Starting this month, the National Research Council of Thailand (NRCT) will put proposals from foreigners to conduct research in these areas under extra scrutiny before issuing a research permit.

The aggressive stance is one facet of revised regulations that require foreign scientists—short-term visitors and residents alike—to obtain research permits. NRCT lists four aims: enhancing cooperation, promoting research that furthers development, controlling natural resources, and "stabilizing the social and economic security of Thailand." Censorship is not a goal, according to NRCT Secretary General Ahnond Bunyaratvej, who has said his office wants to be a research "facilitator, not an inspector."

Off limits? Under new regulations, foreigners may need special clearance to do research on Thailand's seamy street life.

CREDIT: STEPHEN SHAVER/AFP/GETTY IMAGES

NRCT's intention to apply the rules to all foreign-born academics at Thai institutions has touched off a firestorm of criticism. "There is nothing 'facilitating' about these regulations. At best, they imply needless, thick layers of red tape. At worst, they are discriminatory harassment of foreign nationals," fumes one foreigner based at a Thai university who asked to remain anonymous. "Most researchers and university officials here agree that foreign researchers should not be regulated or judged any differently than Thai researchers," adds Warren Brockelman, a conservation biologist who has been teaching at Mahidol University in Bangkok since 1973.

According to the rules, "research involving a foreigner must be conducted jointly with a Thai researcher or consultant." Applications and reference letters must be submitted at least 90 days before a project is slated to start. Lecturers who do not conduct research need not apply for a permit. NRCT will disseminate the rules at a public meeting next week.

In some respects, the regulations codify common sense—and fairness, says Heng Thung, a specialist in satellite data at the Regional Centre for Archaeology and Fine Arts in Bangkok: Scientists who parachute in without knowledge of Thai language or customs should have a local partner. Thung says that many foreign scientists have ignored a long-standing NRCT requirement to deliver a report or thesis after completing a project. Thung also feels that NRCT is justified in taking a stand against foreigners seeking to profit from Thailand's natural resources—developing a drug from a native plant, for instance—without repatriating a portion of the earnings. "Many researchers mine the country," he says.

Thailand also hopes to keep its guard up against questionable research. Earlier this autumn, Korean and Thai newspapers reported that the discredited South Korean stem cell researcher Hwang Woo Suk intends to set up a lab in Thailand. In response, Thailand's science minister, Yongyuth Yuthavong, says he instructed staff members "to be extra-careful about collaborations which raise ethical concerns." NRCT says that it has not received a proposal from Hwang.

What troubles some observers is that NRCT is assuming the mantle of moral arbiter. "Sensitive research will be considered project by project," says Pannee Panyawattanaporn, chief of NRCT's foreign researcher management division. The council will consider factors such as objective, methodology, and the research site, she says. "We want to know whether a project might affect Thailand ▶

"Scientific Deficiencies" at FDA

Twelve months after U.S. Food and Drug Administration Commissioner Andrew von Eschenbach asked for an assessment of FDA's science, the results are in: It's dismal. In a 300-page report by agency advisers and outsiders, 33 experts from industry, academia, and government conclude that FDA is suffering from "serious scientific deficiencies." Those include an inability to adequately monitor the food supply and medical products.

"It certainly sounds consistent" with previous analyses of FDA, says epidemiologist Bruce Psaty of the University of Washington, Seattle, who has criticized the agency's drug-safety record. But the new report released this week is striking for its breadth, says Psaty, covering the vast range of responsibilities at the agency. Among the recommendations: Beef up the agency's work force and form an "incubator" that, among other things, could help define personalized medicine. "This is the science that FDA really has to take the lead on," says Eve Slater, a senior vice president at Pfizer, who helped create the report.

–JENNIFER COUZIN

"Endangered" Rulings Reversed

Seven decisions on endangered species made on the watch of a controversial political appointee should be revised, the U.S. Fish and Wildlife Service (FWS) has decided. Julie MacDonald, former deputy assistant secretary for Fish and Wildlife and Parks, resigned in May after the Department of the Interior's (DOI's) inspector general found that she had inappropriately "reshaped" the science behind decisions related to the Endangered Species Act (*Science*, 6 April, p. 37). House Natural Resources Committee Chair Nick Rahall (D–WV) then asked DOI to review those actions.

FWS has concluded that seven of eight decisions it recently reviewed should be revised, including the choice not to consider listing the white-tailed prairie dog as an endangered species and to withdraw the Preble's meadow jumping mouse from the list. FWS will also reexamine critical habitat designations for the Canada lynx, a toad, a frog, and 12 species of Hawaiian flies. In a statement, Rahall now says the turnaround leads him to "question the integrity of the entire program under [MacDonald's] watch." Others have already come to that conclusion; last month, the Center for Biological Diversity, an advocacy group based in Tucson, Arizona, sued FWS over decisions involving 55 other species.

–ERIK STOKSTAD

negatively," adds another NRCT official, who declined to give his name. The council will consult security officials on sensitive projects, he says.

NRCT's policing role could put it in conflict with the local employers of foreign scientists. The council "should reject a project only after consultation with the host institution" and with the institution's assent, says Brockelman. He notes that the grounds for possible rejection are "rather vague." In a meeting with foreign staff members at Mahidol last week, university officials said that they do not recognize NRCT's right to regulate university staff, according to an attendee. Likewise, Thailand's National Science and Technology Development Agency (NSTDA) "is constructively discussing with NRCT how this measure could be implemented," says assistant president Sirirurg Songsivilai. "We would ensure that NSTDA's overseas researchers and collaborators are not negatively affected by the regulations."

The regulations could be revised after a period of use, says Pannee. In the meantime, foreign researchers have urged NRCT to consider other approaches to facilitating cooperation, such as seminars and lectures. And the council has some fence-mending to do. "Numerous foreign-born academics have devoted their careers to teaching Thai students and involving them in research. Instead of a thank you, they are suddenly told that all of their research is suspicious," says the foreign university scientist, who fears that the rules will put a chill on partnerships.

–RICHARD STONE

BIOTERRORISM

Panel Provides Peer Review of Intelligence Research

In 2002, U.S. intelligence officials claimed that the Iraqi government owned a number of mobile labs capable of producing biological weapons. After U.S.–led forces invaded Iraq, the labs were revealed to be production facilities for hydrogen used to fill weather balloons. In 2005, a government commission said the error was due to a lack of scientific expertise within the U.S. intelligence community.

Now intelligence officials are addressing that problem by opening up their biological research program—most of which is classified—to external peer review. A panel of life scientists from universities, companies, and nongovernmental organizations has begun to assess the merit of projects proposed and conducted by researchers at the 16 agencies under the aegis of the Director of National Intelligence (DNI), as well as grant applications submitted to the agencies. The Biological Sciences Expert Group (BSEG), with 24 core members and an extended network of 40 others, has already met five times this year at DNI's National Counterproliferation Center in McLean, Virginia. In addition to helping screen and design projects to combat bioterrorism, the group will analyze research findings, review the scientific validity of intelligence assessments, and occasionally conduct its own studies.

The objective is to raise the review of intelligence research to the standards of other federal science agencies, says Lawrence Kerr, senior bio adviser at the center and a microbiologist formerly at Vanderbilt University in Nashville, Tennessee, who assembled the expert group. He says the research done by intelligence agencies is currently reviewed primarily by the agencies' own scientists and program managers, who have a limited range of expertise. The system "isn't what one would think of as being incredibly robust," he says.

"Such outreach ought to be standard practice, particularly in fields where rapid changes are taking place," says Dennis Gormley, a senior fellow at the Monterey Institute's Center for Nonproliferation Studies in Washington, D.C., who applauds Kerr for creating the new panel. Kerr says DNI plans to extend the concept to other areas of intelligence research.

BSEG's members are being paid annual retainers of as much as $1000, and its core panelists can earn as much as $18,000 a year based on the amount of work they perform. But DNI has not released their names, part of what Kerr says is a necessary veil of secrecy both to protect them from being snooped on by foreign intelligence agents and to avoid jeopardizing their ties to other scientists. Members even keep two sets of notes at meetings, Kerr says, recording "all of their classified stuff on blue paper or pink paper that's kept separate."

Second look. Lawrence Kerr says outside reviewers will strengthen classified biological research by intelligence agencies.

CREDIT (INSET): COURTESY OF DNI

Some observers are worried that the secrecy surrounding the panel could cloak work on the development of biological weapons. "Remember, this group is not just advising, it may also conduct research," says Alan Pearson of the Center for Arms Control and Non-Proliferation in Washington, D.C., adding that there's a history of similar groups in the past "transition[ing] from defensive to offensive work, rationalizing themselves along the way."

Gerald Epstein, a biosecurity expert at the Center for Strategic and International Studies in Washington, D.C., and the only group member to have made his affiliation public, discounts those concerns. He says BSEG provides an extra layer of oversight to the intelligence community's classified programs and can help the United States remain faithful to the Biological Weapons Convention, to which it and 158 other nations are parties. "If members discovered research that was illegitimate, they could take a number of steps to stop it, such as notifying Congress or even going to the press," he says. Pearson replies that the panel would offer a more credible safeguard "if the broader scientific community knows who they are."

One member of BioChem 20/20, which was formed by the Defense Intelligence Agency in the late 1990s to provide programmatic and strategic advice, thinks that BSEG members will eventually disclose their identities. Geneticist Stephen Johnston of Arizona State University in Tempe says that BioChem members initially chose to remain anonymous. But once satisfied that the work was ethical and noncontroversial, he says, "many of us put the affiliation on our résumés."

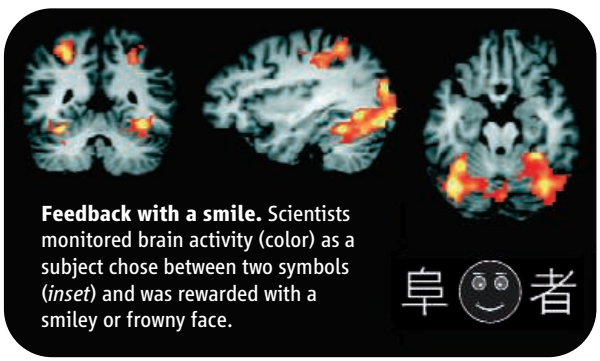
–YUDHIJIT BHATTACHARJEE

NEUROSCIENCE

Gene Variant May Influence How People Learn From Their Mistakes

"Once burned, twice shy" works for most people. But some people are slow to learn from bad experiences. Now, a team of neuroscientists in Germany reports on page 1642 that people with a particular gene variant have more difficulty learning via negative reinforcement.

The research, which combined brain imaging with a task in which participants chose between symbols on a computer screen, centers on the A1 variant, or allele, of the gene encoding the D2 receptor, a protein on the surface of brain cells activated by the neurotransmitter dopamine. Earlier studies have hinted that this variant alters the brain's reward pathways and thereby makes people more vulnerable to addictions.



Feedback with a smile. Scientists monitored brain activity (color) as a subject chose between two symbols (*inset*) and was rewarded with a smiley or frowny face.

CREDIT: T. A. KLEIN *ET AL.*, *SCIENCE*

The new report, from Tilmann Klein of the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany, and colleagues, has earned a mixed reception. Among those impressed is geneticist Bert Vogelstein of the Johns Hopkins University School of Medicine in Baltimore, Maryland, who says that "demonstrating that a single-base-pair difference in the genome is associated with a remarkably different ability to learn from past mistakes is quite an accomplishment."

Klein's team enlisted 26 healthy German males, 12 of them with at least one A1 allele. While undergoing functional magnetic resonance imaging (fMRI), the men performed a learning task that involved looking at three pairs of Chinese ideograms and determining which in each pair was the "good" symbol. For one pair, for example, choosing the good symbol elicited a smiley face 80% of the time; the other times the good symbol was chosen, it elicited a frown. For the other two pairs, choosing the good symbol produced positive reinforcement 60% or 70% of the time. The volunteers viewed each pair 140 times during the learning phase, and the researchers at the end saw no significant difference between men with or without A1 alleles in how well they learned to select the good symbols.

Then, the researchers presented the subjects with the six symbols in various new pair combinations and evaluated how well each man had learned to identify good symbols versus how well they had learned to steer clear of a "bad" symbol. The ones with the A1 allele did a significantly poorer job of not choosing a bad symbol, suggesting they have a deficit in "avoidance learning."

During the initial learning phase, the fMRI scans of subjects with A1 alleles showed less activity in an area of the frontal cortex and the hippocampus—locales involved in negative feedback monitoring and memory—than did those of the controls. A single A1 allele is associated with as much as a 30% reduction in D2 receptor density and means that "the monitoring system seems to respond less to negative feedback," says co-author Markus Ullsperger. He suggests that this phenomenon could be related to impaired reward systems in addictions.

The D2 story remains tangled, however. "Everyone realizes [the D2 receptor] is critical for reward and many other behaviors," says David Goldman of the National Institute on Alcohol Abuse and Alcoholism in Bethesda, Maryland. But, he says, the A1 allele has not been shown to alter how the receptor operates. Geneticist Neil Risch of the University of California, San Francisco, adds that this allele "has been a candidate gene for every imaginable psychiatric phenotype for 18 years now, and to my knowledge none of the originally reported associations has held up."

Nonetheless, cognitive neuroscientist Michael Frank of the University of Arizona in Tucson says the study shows that differences in responses to negative feedback can be "reliably predicted by genetic factors controlling dopamine D2 receptor density" and that this connection is backed up by relevant patterns of brain activation.

–CONSTANCE HOLDEN

Lab Project Launched

A massive biomedical research facility in London received a green light this week when British Prime Minister Gordon Brown announced that one of his government's departments had agreed to sell a key 1-hectare plot to a coalition composed of the U.K. Medical Research Council (MRC), two medical charities—the Wellcome Trust and Cancer Research UK—and University College London. The coveted downtown site was "absolutely critical" for the planned UK Centre for Medical Research and Innovation, says Mark Walport, director of the Wellcome Trust, which will contribute about £100 million to the estimated £500 million project.

Cancer Research UK would transfer its London Research Institute to the new center. And MRC would relocate the government's largest biomedical research outfit—the National Institute for Medical Research (NIMR)—to the site, which may end a long battle over its future (*Science*, 20 April, p. 353).

Paul Nurse, president of Rockefeller University in New York City, will lead a committee charged with developing the scientific plans for the center. Local residents seeking affordable housing for the site may try to block the project, however, as may those who fear that the new labs, which will include NIMR's World Influenza Centre, could expose London to biosafety risks.

–JOHN TRAVIS

Lab Review Panned

The U.S. National Institutes of Health (NIH) did a shoddy job in reviewing the risks of a controversial high-security biosafety lab being built in Boston, according to a new report from a National Research Council (NRC) panel.

The $178 million Boston University lab in the city's densely populated South End will include biosafety level 4 (BSL-4) facilities for studying the deadliest pathogens, such as Ebola virus. In response to a request from the state of Massachusetts, NRC reviewed an NIH draft study of alternative sites outside Boston and of worst-case scenarios if a pathogen escaped (*Science*, 11 August 2006, p. 747). The NRC panel labeled the study "not sound and credible," faulting NIH's modeling and its failure to consider the escape of highly transmissible agents such as avian influenza and SARS. Construction on the lab will continue as NIH sifts through the NRC report and other comments. But the pending risk assessment could delay resolution of federal and state lawsuits opposing the operation of its BSL-4 suites.

–JOCELYN KAISER

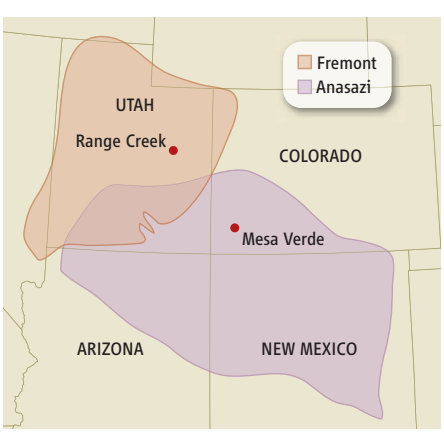
The Vanishing Fremont

What forced the Fremont Indians into sky-high cliff dwellings 1000 years ago, and why did they disappear a few hundred years later? A rugged Utah canyon yields new clues

WHEN LARRY COATS PULLED HIMSELF up to an overhang near the top of a pinnacle in Range Creek Canyon, more than 200 meters above the valley floor, his left foot landed on a perfect square that had been pecked into the rock face, exactly where a climber would want it. He looked up and a mysterious, hidden world from Utah's prehistoric past revealed itself. An obscure buttress held pictographs only visible from this precarious spot. Climbing even higher, he followed a series of steps carved into the sandstone wall that led to the multipinnacled summit.

From there, Coats, a paleoecologist at the University of Utah in Salt Lake City and a professional climber, rappelled down the rock face to explore lower ledges and overhangs that were accessible only from the summit. He found 1000-year-old granaries—food storage caches made of mud, stone, and wooden poles—built into the cliff wall, as well as pottery remnants, more rock art, the outlines of subterranean pit-house structures, and a metate, a stone used to grind corn. The metate "was perfectly placed, tucked carefully under a ledge, as though someone was intending to come right back and get to work grinding maize," recalls Coats, who has surveyed the cliff's pinnacles over the past two summers with archaeologists.

The sky-high lodgings and accouterments that Coats found are part of a much larger constellation of sites currently being documented in this remote canyon nestled behind the towering Book Cliffs, 240 km southeast of Salt Lake City. Range Creek's reclusive owner sold it to the state of Utah in 2001, and archaeologists have been amazed by its spectacularly preserved ruins. Newly dated to roughly 1050 C.E., Range Creek was one of the more populous settlements of the Fremont people, enigmatic farmer-foragers who lived mostly in what is present-day Utah and western Colorado.



Under the same sun. The neighboring Fremont and Anasazi peoples faced severe ancient droughts.

After years of surveying, researchers have begun to work the giant site in earnest. They made their first round of excavations last summer and are working to build a tree-ring record and gather radiocarbon dates, says archaeologist Duncan Metcalfe, who heads the Range Creek Research Project at the University of Utah, Salt Lake City. The scientists hope that the site's archaeological riches, including an apparent network of caved-in pit-house villages just above the valley floor, will yield insight into what is perhaps the greatest mystery concerning the Fremont: Why did they vanish?

Metcalfe and others believe that the conditions that led to the Fremont's puzzling disappearance in 1300 C.E. are vividly expressed in Range Creek's social

CREDIT: LARRY COATS/RANGE CREEK PROJECT; SOURCE: JESSE JENNINGS, *UNIV. OF UTAH ANTHROPOLOGICAL PAPERS* #98 (1978)

High art. Only professional climbers can get a close view of many of the Fremont pictographs and artifacts left in the cliffs of Range Creek.

disintegration roughly 150 years earlier. "Range Creek is like finding a new library vault full of information," says Kevin Jones, Utah's state archaeologist and a member of the Range Creek research team. "Those books are going to be extraordinarily telling and valuable."

Indeed, Metcalfe and others believe that Range Creek's secrets may ultimately reveal information beyond the Fremont culture itself. The Fremont's sudden collapse 700 years ago parallels that of other long-standing Southwestern cultures, including the Anasazi, the Fremont's cliff-dwelling neighbors in the Four Corners region. Experts consider the Southwest in the 1200s to have been extremely tumultuous but are split over which was the greater destabilizing force: a downturn in the environment that made farming untenable or a fracture in the social order. Whatever the trigger, fear and violence seem to have spread like wildfire throughout the region in the 13th century. At Mesa Verde in Colorado, for example, the Anasazi sought shelter high in the cliffs and left abundant evidence of gruesome violence and cannibalism (*Science*, 8 September 2000, p. 1663).

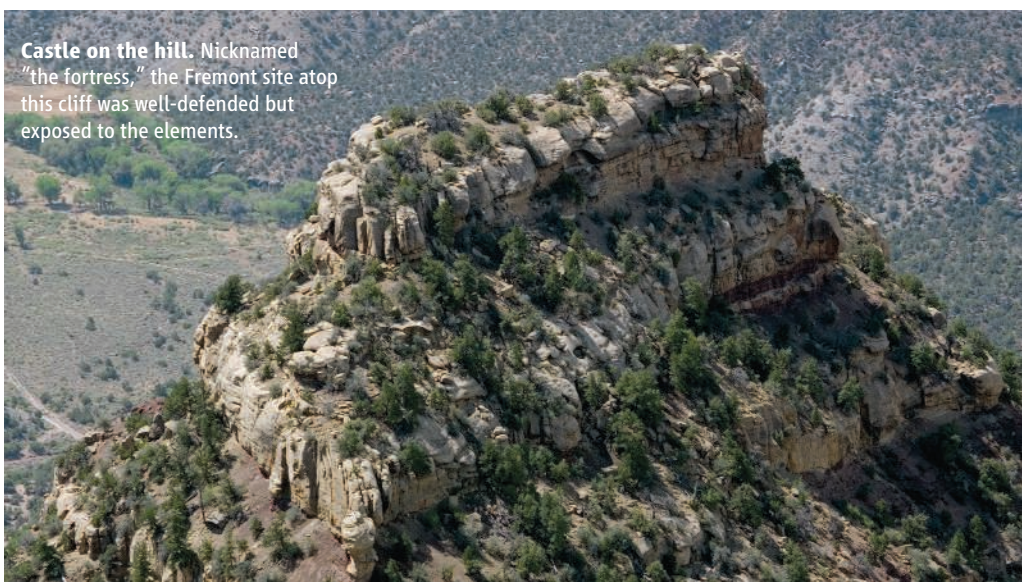
The social upheaval extends south, reaching into places like present-day Phoenix, where the Hohokam and other prehistoric peoples also massed together in self-defense in the 13th century. "By 1275, everybody in the Southwest is living in a fort," says archaeologist Steven LeBlanc of Harvard's Peabody Museum of Archaeology and Ethnology, whose numerous works assert that the Southwest at the time was beset by warfare.

Range Creek, where the Fremont uncharacteristically clustered in fortresslike hideaways and stashed their food in hidden caches, suggests similar strife some 100 or 150 years earlier. Located at the extreme northern fringe of the Southwest, Range Creek was always marginal for farming, notes Metcalfe. Climate changes and subsequent new survival strategies may have occurred there "in advance of what's happening later throughout the Southwest," he says.

Farmers, foragers, or both?

In the late 1920s, young archaeologist Noel Morss was exploring central Utah's rugged canyon country when he found gray pottery, moccasins constructed from deer hocks, and visually arresting trapezoidal figures, which were displayed on clay figurines and pictograph and petroglyph panels along the banks of the Fremont river. These material traits had previously been considered an offshoot of an earlier farming culture that flourished south of the Colorado River from 750 C.E. to 1300 C.E. and is known today as the Anasazi or Ancestral Puebloans. But Morss felt the artifacts showed an originality that set them apart from the Anasazi, who wore sandals, lived in elaborate cliff dwellings, and drew Kokopelli and stick figures. He concluded that the "Fremont drainage proved to be the seat of a distinctive culture." His general characterization of the Fremont has held up remarkably well, and the name stuck.

Ensuing excavations of Fremont sites throughout Utah have uncovered a hodgepodge of hunting, farming, and foraging habits, however, and today the Fremont are rather fuzzily defined. One camp maintains that they were country cousins of the Anasazi, primarily farmers living in pit houses. Another camp has contended that they developed in situ from a preestablished archaic culture and were predominantly hunter-gatherers who incorporated farming into their repertoire. "We still don't know who they were, much less what happened to them," says Fremont expert Jerry Spangler, executive director of the Colorado Plateau Archaeological Alliance, a Utah-based antiquities preservation group. He and some others define the Fremont as "farmer-foragers" who switched between hunting and gathering and farming, depending on circumstances. That versatility makes their collapse even more of a mystery. "They could do it all," says Spangler. "That's what makes them so unique among southwestern cultures."



Castle on the hill. Nicknamed "the fortress," the Fremont site atop this cliff was well-defended but exposed to the elements.

Range Creek turns out to be an ideal laboratory for discovering more about the Fremont. The rugged landscape, which today remains sparsely inhabited and virtually roadless, has helped keep both the archaeology and the ecosystem largely intact for the past 1000 years. That makes the site a rare prize for southwestern archaeologists, who often work only one step ahead of developers breaking ground for roads and houses. "This is the first time in my life I won't have a bulldozer at my back," says Metcalfe.

Range Creek's protection should continue, because the University of Utah has recently secured a 20-year lease on a 486-hectare parcel containing the greatest concentration of Fremont ruins. Metcalfe runs a summer field school there as well as a multidisciplinary platoon of scientists, with grants totaling roughly $300,000 per year from the state of Utah and the National Science Foundation.

But the work is slow: Metcalfe's team has surveyed just 10% of the 20,234-hectare canyon and expects the total number of sites to number in the thousands. And although researchers have spotted five sets of human remains eroding out naturally, Metcalfe says he has no plans to disturb them, much less do DNA testing. Because Range Creek is largely public land, soon after archaeologists discovered the bodies they notified neighboring Indian tribes as required by law. Tribes typically prefer human remains to stay in their original place of burial. To prevent any conflict, researchers find it easier to avoid the bones altogether.

Instead, Metcalfe and others focus on the granaries and pit houses. Together, these artifacts suggest a society under stress, competing for dwindling food sources, and splintering

CREDIT: LARRY COATS/RANGE CREEK PROJECT

Safe storage. At Range Creek, the Fremont cached their food in hard-to-reach granaries such as this one (*inset*).

into self-protective encampments. Of the nearly 400 sites documented in the canyon thus far, 80 are granaries placed far up cliffs and concealed on narrow ledges or under overhangs. The dense concentration of these storage chambers and their hidden nature is "unprecedented" in Fremont history, says Spangler. Range Creek's granaries "are the most inaccessible I've ever seen," agrees Coats, even harder to reach than those of the Anasazi, which were also generally made of mud and stone and perched on cliff ledges. The inaccessible food caches suggest that the Fremont were defending their food supply. Jones speculates that the Fremont were "scatter hoarding," spreading their food out in multiple hiding places: "You risk losing some of it, but at least if another person gets into it, they've only got one bit."

Last summer, Coats found additional compelling evidence of defensive settlements. For example, above a pit-house site researchers had dubbed the "deluxe apartment in the sky" because it is nearly 300 meters above the valley floor, Coats found piles of boulders strategically placed at the access points of the ridge. There's even a log still wedged underneath one of the big rocks. "I assume a lever was in place at one time, where they could release the rocks down onto anyone who was approaching," says Coats. "It certainly looks like a defensive weapon."

Another site—atop a butte and "exposed to all the weather," says Coats—is nicknamed "the fortress" because it contains similar walls of boulders perched at the edges of the ridge. Here Coats observed numerous artifacts on the ground by four well-used pit houses, including metates, pottery fragments, and lithic flakes. All this "indicates quite a lot of activity on top for a significant amount of time," he says, and suggests long-term, rather than seasonal, occupation. Researchers have discovered numerous other dwellings wedged on the tops of steep ridges, although the Fremont's cornfields were apparently far below in the floodplain. These houses "are not next to their farm fields, and they are places where grandma and grandpa would have a hard time getting to, and where your children, with one misstep, would fall and get hurt or die," says Jones. "Why would you live in a place like that?"

For defense and safety, researchers hypothesize—but defense against whom? Until recently, Jones and other Range Creek researchers thought that the cliff-top dwellings represented the terminal stage of the Fremont, in the 1200s or 1300s, the same time that the Anasazi retreated into cliff dwellings at Mesa Verde. Jones had expected that the lower sites just above the valley floor, where huge circular stone alignments suggest pit-house villages, were earlier, perhaps 900 C.E. to 1100 C.E.

But the new data reveal that the Fremont on the ridge tops and in the villages may have co-existed, perhaps about 1050 C.E. A dozen radiocarbon dates, obtained from corn, arrow shafts, granary beams, pit-house rafters, and other organic material have produced a tantalizing pattern, says Metcalfe. Of the 12 dates, 10 share a 95% confidence interval that falls between 970 C.E. and 1130 C.E., with the average falling at 1050 C.E. To further narrow the range, Metcalfe plans to use tree rings, which offer accurate dating to the year.

Based on the density of pit-house alignments in the valley, Metcalfe estimates that a total of about 1000 Fremont lived in the canyon. But there are no trash middens, as expected if the Fremont had a long-term presence there, he says. "They stayed for a relatively short time and got out fast," he believes.

The defensive settlements and 1050 C.E. date are commensurate with sites in nearby canyons along the Green River, including Nine Mile, a spectacular rock-art site that also features remote granaries and fortresslike structures atop ridge tops. "Everything we have in Nine Mile and Range Creek points to groups of people protecting themselves and their food," says Spangler. In addition to the shields and human combat depicted in Nine Mile's rock art panels, researchers in 1992 found a child buried with an arrow point in its chest cavity.

The Fremont also apparently massed together later, during the 1200s, on a scale much larger than Range Creek at a site called Five Finger Ridge in south-central Utah. The giant site is radiocarbon dated to between 1200 C.E. and 1300 C.E., near the end of the Fremont period, and includes remote granaries tucked high in the cliffs and more than 60 structures, including pit houses, packed tightly together on a knoll.

At Range Creek, if the dates for valley dwellings do indeed coincide with those in the ridges, it's unlikely that the Fremont were protecting themselves from outsiders: The whole region was settled by Fremont, says archaeologist Joel Boomgarden, a member of the research team. "I'd be willing to bet it's from people within the canyon. They're probably defending themselves against their neighbors."

The dry years

But why? Was it social or climatic factors, or some combination of the two, that splintered Fremont society? New climatic records offer clues. In several studies published this year, paleoclimatologist Larry Benson of the U.S. Geological Survey (USGS) in Boulder, Colorado, plotted out a series of major droughts that pummeled the Midwest and Western regions from the early 11th century to the end of the 13th century. He borrowed a drought index—which uses prehistoric tree-ring data on precipitation and temperature to estimate soil moisture—from Edward Cook of Columbia University and his colleagues. The index charts conditions year by year, which Benson and colleagues then compared

CREDITS (TOP TO BOTTOM): LARRY COATS/RANGE CREEK PROJECT

to events in some of the agrarian cultures that melted away during this span, such as the Anasazi, Fremont, and Cahokia; the latter farmed the Mississippi River floodplains and valleys.

If the Range Creek occupation was in fact at its height about 1050 C.E., it coincides with one of a series of decadal-long droughts in the region, says Benson. "There is a 20-year drought in the Four Corners [area] centered at 1050, and it follows a pretty dry period that lasted much longer," he says.

The match between climate and cultural upheaval becomes even clearer in the next 2 centuries. Benson notes that scientists now consider the mid–12th century megadrought (1135–1180 C.E.) to be the most severe in the past 2000 years. At this time, the Anasazi abandoned their main hub in Chaco Canyon in present-day New Mexico and started bunching together in Mesa Verde's cliffs in Colorado. There, "average precipitation during this drought was reduced by 11%, with some years seeing a reduction in the mean of approximately 50%," says Benson.

A century later, at about the time the next persistent drought (1276–1299 C.E.) is over, both the Fremont and Anasazi are gone from their ancestral homelands. "In some sense, the 13th century drought may have simply 'finished off' some cultures that were already in decline," Benson and his co-authors wrote earlier this year in *Quaternary Science Reviews*.

Those droughts went beyond the Southwest, impacting much of the contiguous United States, he says. A close reading of the drought index shows that the mid–12th century drought "was impacting the Midwest, from Illinois, all the way to the coast of California," says Benson. "The climate is causing crops to fail in the Four Corners where the Anasazi were based, and in Utah, where the Fremont lived; and it is probably also causing crops to fail in the Mississippi valley."

The impact was all the harder because of the previous and intervening wet years, researchers suspect. There's evidence that after each drought the Fremont rebounded as climate improved. "Each time they did that, there seems to be a population boom," says Steven Simms, a Fremont scholar and archaeologist at Utah State University in Logan. But those extra mouths to feed demanded more crops, leaving the culture even more vulnerable to the next dry spell.

Other climate change forces may also have been at work, including a cold period in the 900s and in the late 1100s suggested by a new analysis of pollen data, says archaeologist Timothy Kohler of Washington State University in Pullman. For the Fremont, eking out a living in an environment already marginal for agriculture, an earlier frost and shorter growing season would have been yet another major hurdle.

To some, all this adds up to a persuasive case for climate change. "I think the evidence for drought as a forcing mechanism is starting to get pretty obvious," says archaeologist Michael Berry of the U.S. Bureau of Reclamation in Salt Lake City, one of Benson's co-authors. "It's not just a factor. It's a forcing factor."

If a deteriorating climate triggered food shortages, some researchers speculate that social disorder resulted. "It starts to tear at the social fabric," says Simms. "That's why late in the record, like in Range Creek, you see these granaries perched way up in the cliffs. It's very much like when you get an oil embargo and you get fistfights at gas stations."

Other experts agree and say that this kind of climate-triggered chaos may have reverberated on a larger scale across the Southwest in the 1200s. "If the crops aren't working, you might start blaming your rain priests or your ideology," says Jeffery Clark, an archaeologist at the Center for Desert Archaeology in Tucson, Arizona.

But the Fremont were such generalists, switching from farming to foraging so easily, that other researchers argue that the megadroughts alone wouldn't have killed them off. "It had to be a combination of circumstances that caused this culture to end after 1000 years of success," says Simms. "It can't just be drought. They weathered those before."

That's why many southwestern archaeologists favor a mix of environmental and social causes. "Environment is always a factor," says archaeologist Carla Van West of the SRI Foundation, a New Mexico–based historic preservation organization. "The question is whether it is a causal, proximate, or an ultimate cause."

Archaeologist Christy Turner of Arizona State University in Tempe has hypothesized that brutal social and political control, a kind of religious terror, was exported into the American Southwest from Mexico about 900 C.E., when evidence of cannibalism starts to show up in the Four Corners region (*Science*, 1 August 1997, p. 635). Turner speculates that the practice spread like a virus and eventually caused populations to splinter and coalesce in defense, until they eventually collapsed.

Yet other scientists find the climatic evidence hard to argue against. Says Julio Betancourt, a USGS paleoclimatologist based in Tucson, Arizona: "If you have three or four corn crops in a row failing, they're going to be a dead people; they're going to starve to death. You can bring culture all you want into the picture; it's not going to matter."

That the Cahokia's great mound-building culture in the Midwest collapsed at the same time as the Anasazi and Fremont strikes some as beyond coincidence. Says Boomgarden: "It almost seems like the link has to be climate, because populations that far apart shouldn't have much to do with each other."

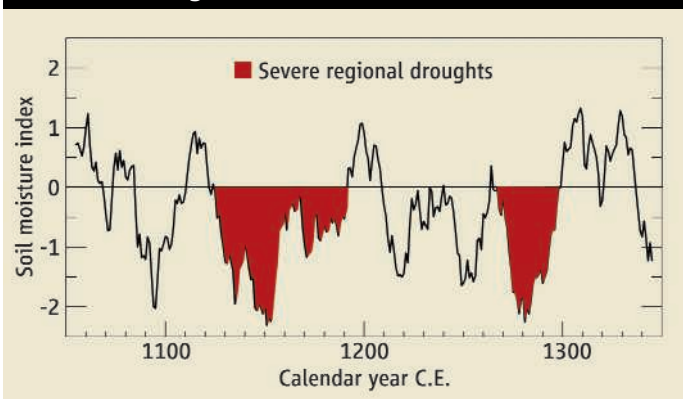
As research on these cultures continues, Range Creek, because of its archaeological and ecological purity, is expected to provide a crucial piece of the puzzle. Next summer, Coats wants to scale the ridgeline across the canyon to search for additional cliff-top dwellings. "I'm convinced there will be more sites up there," he says. By then, Metcalfe, who plans to expand excavations to include several sky-high sites, hopes to have the tree-ring history of Range Creek in hand, revealing precisely when the drought struck the canyon.

Still, exactly why the Fremont and other cultures sought refuge in the cliffs may elude researchers for some time. Says Spangler: "Is it warfare for warfare's sake? Is it warfare because of environmental stress because you can't produce enough food? There are multiple lines of evidence for each argument."

–KEITH KLOOR

Keith Kloor is a senior editor at *Audubon* magazine.

Ancient Droughts in the Four Corners Area



CREDIT: LARRY BENSON/USGS

Impossible dream? At an October meeting, Bill and Melinda Gates challenged the world to eradicate malaria in their lifetimes.

MALARIA

Did They Really Say . . . Eradication?

The malaria world is all abuzz about a call by Bill and Melinda Gates to wipe the scourge from the planet. Even if it proves unfeasible, their idea could have a big impact

SEATTLE, WASHINGTON—When Bill and Melinda Gates had finished their back-to-back speeches, many researchers could barely believe what they had just heard. At a meeting hosted by their charitable foundation in their hometown, the couple had uttered the long-forgotten e-word, calling for a sweeping new plan to eradicate malaria.

At first, some thought the philanthropists had misspoken. Very few people have talked about eradicating malaria since an earlier program crashed and burned in the 1960s, leaving a permanent smudge on the field and resulting in a resurgent epidemic across much of the globe. Malaria now kills more than a million people a year, and some malaria experts say eradication, although a noble goal, is simply unachievable. Yet the speeches delivered at the Gates Foundation Malaria Forum on 16 to 18 October leave no room for doubt: The couple wants the malaria parasite to go the way of the smallpox virus.

The second surprise came after the speeches, when Margaret Chan, director-general of the World Health Organization (WHO) in Geneva, Switzerland, jumped up, grabbed a microphone, and enthusiastically seconded the idea. "I pledge WHO's commitment to move forward, and I dare you all to come along with us," she said, reportedly without consulting some of her senior lieutenants.

Chan and the Gateses were careful not to set a target deadline, presenting eradication as a long-term vision, not a near-term goal. "Multiple decades" is what Bill Gates told reporters afterward, noting that it is "dangerous" to offer anything more concrete. "They both hope it will happen in their lifetimes," says Regina Rabinovich, head of infectious diseases at the Gates Foundation, who is intimately involved with the plan. Even with those caveats, the call has ignited a debate on whether it is wise—given a long history of broken health promises—to dangle potentially unattainable goals before the public. "There is a danger of overpromising and underachieving," says Joel Breman, senior scientific advisor at the Fogarty International Center at the U.S. National Institutes of Health in Bethesda, Maryland.

But at the same time, the daring call is having a major impact. Bolstered by already-plummeting malaria rates in several countries, a group of informal advisers has formed a kitchen cabinet of sorts, loosely composed of heavyweight scientists and senior officials from the big funding agencies in malaria, to try to turn the lofty vision into reality—or at least see how far they can get. "It has galvanized the community and created quite extraordinary momentum," says Rajat Gupta, chair of the board of the Global Fund to Fight AIDS, Tuberculosis, and Malaria, who is a member of that group.

The Roll Back Malaria (RBM) Partnership, composed of all the major players in malaria, including the endemic countries, has already lent its support. Meeting in the Ethiopian capital, Addis Abeba, last week, the RBM Board agreed to set up a high-level steering committee to coordinate efforts and devise a "business plan" within 6 months. No new funding has been announced, but everyone expects the Gateses to put large sums of money where their mouths are.

Reality check

In the wake of the Seattle meeting, proponents have been trying to reassure skeptical scientists and manage expectations, in part by de-emphasizing the importance of words. Scientists use "eradication" to mean that a pathogen no longer exists anywhere on Earth—save for perhaps a few lab freezers—and control measures can stop. "Elimination" means a pathogen is no longer transmitted in a defined geographical area, although "imported" cases may still occur. By those definitions, malaria has been eliminated in Europe, measles in the Americas, and polio in most countries of the world—but smallpox remains the only disease that has been eradicated.

"I like the term 'elimination' better" than eradication, Chan told *Science* in Seattle, shortly after the Gateses issued their call. "Eradication is of course the ultimate goal, and I don't mind people using [the words] interchangeably. … It is elimination-slash-eradication, depending on the availability of tools."

Theoretically, there's little doubt that malaria could be eradicated, because there's no animal reservoir from which the disease could bounce back into the human population after it's gone. Nicholas White of Mahidol University in Bangkok believes eradication is already within reach using the latest weapons, such as long-lasting insecticide-treated bed nets, powerful new drugs called artemisinin-based combination therapies (ACTs), and indoor insecticide spraying (*Science*, 26 October, pp. 556 and 560). Where these weapons have been mass-introduced, malaria is retreating fast, says White.

But most others, including Bill and Melinda Gates, say that although current methods can eliminate malaria in some areas, they won't suffice for global eradication; more

CREDIT: AP

powerful ways to break the transmission chain are needed in the hardest-hit areas. "We do not have the tools that are needed to complete malaria eradication today," says Rabinovich.

That's one key distinction that sets this initiative apart from the previous failed eradication effort, says Carlos "Kent" Campbell, former head of the malaria branch at the U.S. Centers for Disease Control and Prevention in Atlanta, Georgia, and now at PATH, a Seattle, Washington–based nongovernmental organization, where he directs the Malaria Control and Evaluation Partnership in Africa program. That earlier effort, which was abandoned in the late 1960s, relied on DDT to wipe out the mosquito vector and on chloroquine to treat the disease, only to see the vector and parasite develop resistance to both.

The Gateses outlined a two-part strategy: Go as far as you can with existing tools while simultaneously investing heavily in new ones. The latter would likely include transmission-blocking vaccines and drugs; new, preferably single-dose, drugs to replace ACTs when they inevitably are rendered ineffective by resistance; and alternative insecticides and even nonchemical means to defeat mosquitoes, such as traps or genetic modification, along with rapid diagnostics and monitoring for resistance—none of which exists today. The "beauty of this approach," as opposed to the earlier one, says Campbell, "is that it links a very specific research agenda with a control agenda."

Existing tools would be massively scaled up over the next 3 to 5 years, says Gupta of the Global Fund. The goal, he says, is to "reduce dramatically, or even eliminate, mortality from the disease and reduce the number of new infections to much smaller numbers." The first step, he says, will be to bolster country programs, then to scale up regionally and finally globally. "The regional approach is very, very important. You can't have a great program in the Zambia and no program in the DRC [Democratic Republic of the Congo]. It doesn't work." To pull it off, he predicts that donors such as the Global Fund, the World Bank, and the President's Malaria Initiative will need to roughly triple the money now available for malaria control, up to $3 billion or $4 billion a year. Gupta calls it a "no-regrets policy. … It doesn't matter when the science comes along; let's just control as aggressively as possible."

Although the hardest-hit countries in Africa and elsewhere are the most obvious targets, public health officials should simultaneously start picking "low-hanging fruit," says Richard Feachem, former executive director of the Global Fund and now head of the Global Health Group at the University of California, San Francisco. By that, he means trying to eliminate malaria from the "natural margins" or edges of the endemic zones, where the disease isn't as entrenched. The result would be a gradual "shrinking of the malaria map." Such an effort is getting under way in southern Africa, where the 14 members of the Southern African Development Community have declared their intention to eliminate malaria, starting with the southernmost countries of Botswana, Namibia, South Africa, and Swaziland and moving north, says Feachem. Elimination plans are also afoot for archipelagoes such as Vanuatu and the Solomon Islands.

Building an arsenal. The earlier failed eradication plan relied primarily on DDT to kill mosquitoes; the new initiative would use every tool in the shed, like bed nets—and many that don't yet exist.

CREDITS (TOP TO BOTTOM): WALTER SANDERS/TIME LIFE PICTURES/GETTY IMAGES; GIDEON MENDEL/CORBIS

False hope?

Nobody would argue against any plan that can have a dramatic impact on malaria. Still, whether all these new activities are a prelude to eradication—and whether it's wise to use that term—is under intense debate. Medical entomologist Willem Takken of Wageningen Agricultural University in the Netherlands thinks it's much too early. Recent victories may be more tenuous than some people realize, he says: Already, researchers are seeing an increase in resistance to pyrethroids, an important class of insecticides, in West African mosquitoes. The ACT miracle, too, is bound to fade, and a vaccine has yet to materialize. Talking about eradication now is giving affected countries false hope, he says. Some also worry about what is called the "Gates Effect"—the fact that the Gateses' vast coffers make people reluctant to criticize them or their projects.

At least as important as the push for new tools is a similar investment to improve the weak health infrastructure across Africa, cautions Donald Hopkins, who leads the global Guinea worm eradication effort from the Carter Center in Atlanta. Even with perfect tools, he says, "we would need capability in each village 24/7. That's not there." Perhaps the biggest challenge, even proponents agree, will be to sustain interest and funding over the long haul. "We are having a difficult time keeping polio eradication going, and it's only been 20 years," says WHO Assistant Director-General David Heymann, who oversees that effort. Originally targeted for completion in 2000, the campaign has stalled in a few especially tough countries and is having a hard time raising enough money to finish the job. Feachem, too, agrees that keeping up the commitment will be difficult—and paradoxically, more so as the end nears. "It will require exceptional leadership," he says. "Luckily, Bill and Melinda are young."

Some say that malaria fighters would do better to take a page from the measles book. Without making eradication an official goal, a sweeping campaign against that viral disease has made impressive strides; just last week, WHO announced a 91% drop in African measles deaths since 2000. Public health officials can hope for eradication—and some certainly do—but they don't have to worry about a backlash if the remaining centers of infection turn out to be impossible to mop up.

But others say what's important is to focus on the big picture. "I think there will be good to come out of this even if malaria eradication proves unachievable in our lifetimes," says Hopkins. Adds Chan: "We need champions like Bill and Melinda."

–LESLIE ROBERTS AND MARTIN ENSERINK

Scan man. Paul Tafforeau readies a Neandertal jaw for scanning by the synchrotron.

PALEONTOLOGY

Paleontologists Get X-ray Vision

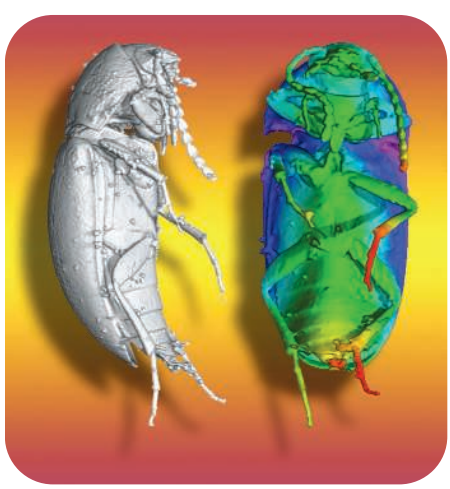
By using x-rays generated from a synchrotron, researchers are getting sharper views of everything from Neandertal teeth to dinosaur embryos

GRENOBLE, FRANCE—On a crisp morning in October, three museum curators crowded into an experimental station on one of the world's most powerful synchrotron particle accelerators. As electrons spun around a giant circular ring nearby and vacuum pumps hummed in the background, the curators painstakingly unveiled the precious fossils they had escorted here from Berlin, Tel Aviv, and Zagreb. "This is my baby, my third son," said Almut Hoffmann, a historian from the Museum for Pre- and Early History in Berlin. She was still a bit wary of handing over the jawbone of a 40,000-year-old teenage Neandertal from Le Moustier, France. "I heard they will not damage it," she said nervously.

French and American researchers spent months convincing Hoffmann and two other curators that it was safe to bring their "babies" to Grenoble to be x-rayed by a beam so powerful it would kill a living human within an hour (and cause cancer-causing mutations within a few seconds). These prehistoric youths lived short, obscure lives, but in death they are much in demand: Daily growth lines in the enamel of their teeth offer an unparalleled record of ancient life history and the evolution of childhood. These lines could show whether Neandertals grew up slowly as modern humans do or more rapidly as apes and early human ancestors did.

A couple of brave curators recently allowed researchers to slice open three Neandertal teeth to read the record of these lines, although the results are conflicting (see sidebar, p. 1547). Now the 52 kilo-electron-volt synchrotron x-ray beam used at the European Synchrotron Radiation Facility (ESRF) allows researchers to detect daily rhythms without cutting or damaging teeth—and so to gather data on many specimens. "Before this technique, the only way we could see this much detail inside was to cut the tooth," says paleoanthropologist Tanya Smith of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. "No other scanner in the world has been adapted to do this."

Grenoble is one of just three large third-generation synchrotrons in the world and the only one adapted for viewing large fossils so closely. So far, it has uncovered new species of ants and beetles trapped in opaque amber, revealed dinosaur embryos encapsulated in eggs, and explored rodent and primate teeth. "It's an entirely new technique for paleontology," says paleontologist Philip Donoghue of the University of Bristol in the U.K., who began using the lower energy Swiss Light Source synchrotron in Switzerland on fossils 3 years ago (*Science*, 13 October 2006, p. 291). "Synchrotron tomography allows us to look at new areas of science without destroying fossils." Other synchrotrons in Europe and China are now gearing up to image fossils, too.



Beetle mania. The synchrotron revealed a Cretaceous beetle entrapped in opaque amber.

The track record of the synchrotron in Grenoble for not damaging fossils finally convinced curators to gather there for 8 days in October, bringing fossils of Neandertals and modern humans that died tens of thousands of years ago in caves in Croatia, France, and Israel.

The paleontologist who made this possible is Paul Tafforeau of ESRF. He started as a graduate student in paleontology at the University of Montpellier in France, where he was unhappy about destroying primate teeth to study their enamel. He began working at ESRF "by accident" after a conversation in 2000 with the head of the imaging group, José Baruchel. He quickly realized the potential of the nondestructive imaging tool, which has four times better resolution than the best conventional computed tomography (CT) scanners that image large fossils. Although his first tests failed, he began working with the imaging group to produce three-dimensional (3D) images of the internal microstructure of primate teeth.

The beauty of the synchrotron is that it produces x-rays that are far more spatially coherent than beams from conventional CT scanners, so the waves of x-rays are tightly in sync when they pass through an object. This coherence is essential for phase-contrast imaging, which allows researchers to see not only how different densities of material in a tooth absorb the x-rays but also how the passing x-ray's wavefront is modified by the electronic structure of the sample. This method exposes even more detail about tooth microstructures smaller than the width of a single cell, Tafforeau and Smith reported online in the *Journal of Human Evolution* on 28 November.

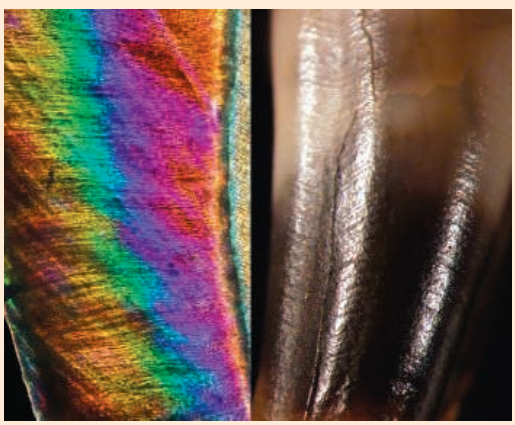
ESRF is also the only synchrotron that can scan larger objects, such as complete

CREDITS (TOP TO BOTTOM): A. GIBBONS/*SCIENCE*; P. TAFFOREAU, R. BOISTEL, M. LAK, A. NEL, AND D. NERAUDEAU

Dental Evidence Suggests Neandertals Matured Faster Than We Do

Paleoanthropologists eager to compare the development of Neandertals and modern humans waited for years to be allowed to take a slice out of a Neandertal's tooth to see the minute daily growth lines inside. "We await a brave curator somewhere who will allow a single Neanderthal tooth to be sectioned; much depends on it," paleoanthropologist B. Holly Smith of the University of Michigan, Ann Arbor, wrote in an article in *Evolutionary Anthropology* in 2004.

Smith has gotten her wish recently, but with mixed results. A study in *Nature* last year of two sliced Neandertal teeth found that the teeth formed slowly, like those of modern humans. But this week in the *Proceedings of the National Academy of Sciences* (*PNAS*), researchers analyzed growth lines in a sliced Neandertal molar plus other uncut teeth from the same specimen. They conclude that this 8-year-old Neandertal from Belgium grew up more rapidly than modern human children, according to lead author Tanya Smith (no relation) of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. "I think it's pretty convincing," says paleoanthropologist Jay Kelley of the University of Illinois, Chicago. But he notes all the same that the paper provides "data for [only] one individual." Data on more Neandertals may be able to resolve the problem this year, thanks to a new method for seeing growth lines without damaging specimens (see main text).



Lifelines. A Neandertal's tooth has both internal (*left*, diagonal lines) and external (*right*, horizontal lines) striations that record its growth.

Researchers have known for some time that humans are the only animals to have extended their childhoods long enough to have a teenage phase. *Homo sapiens* grew up twice as slowly as apes and our australopithecine ancestors that lived 4 million to 2 million years ago, says Holly Smith. Our ancestors may have lengthened childhood and delayed reproduction to allow more time to develop their brains, perhaps improving social learning, language, and other behaviors.

But researchers do not know when this dramatic change in life history strategy took place in the human family. Were *H. erectus* parents 1.8 million years ago the first to experience the joys of teenagers, or did adolescence appear 500,000 years ago in a common ancestor of Neandertals and modern humans? A previous report in *PNAS* by Tanya Smith suggested that it was even later.

The best way to find out is to look inside the teeth of Neandertals, modern humans, and their ancestors. Like rings in a tree, teeth grow incrementally, preserving a record of their development in microscopic lines in their enamel. These lines are deposited daily, along with less frequent lines that reveal stresses such as birth. And longer-period lines can be seen on the surfaces of teeth. In the new study, Tanya Smith took a thin section of an upper molar and counted the number of daily lines laid down before and after birth and between long-period lines. This told her how many days passed between the longer-period lines. She could then use the external lines on the uncut teeth to calculate how much time passed before their roots and cusps formed completely, as well as to determine the timing of key developmental benchmarks. She found, for example, that the second molar erupted a few years earlier in this 8-year-old Neandertal than in *H. sapiens*, suggesting that Neandertals grew up faster than we did.

That conclusion contradicts the earlier study of a Neandertal, done by Christopher Dean of University College London and colleagues. Given the conflicting reports, the next step is to analyze more specimens. "Dental evidence from a larger number of individuals … would go a long way toward clinching the claim that they were distinct in the way they grew up," says Dean. That is precisely what Tanya Smith and her colleagues are trying to do with their new x-ray vision in Grenoble. **–A.G.**

hominid skulls, at a 45-micrometer resolution (the width of a hair). By taking radiographs of a sample that rotates 180° or 360° during a 2-hour run on the beamline, the team can use software to produce a stack of cross sections that generate a precise 3D image. The cost for the 8-day run on hominid teeth: $120,000, in this case underwritten by ESRF.

So far, Tafforeau and colleagues have used the synchrotron to expose the internal structures of fossil green algae and an unerupted premolar in an extinct primate, among other fossils featured in *Applied Physics* in 2006 to demonstrate the method. Detailed new images of dinosaur embryo bones are "truly spectacular and cause a stir every time they are shown at a scientific meeting," says paleontologist Eric Buffetaut of the Centre National de la Recherche Scientifique in Paris.

Tafforeau was recently hired full-time at the synchrotron to focus its x-rays on more fossils. Specimens that recently vied for precious time on the beamline include Cretaceous mammals encased in rock, dinosaur and bird embryos, snails, rodent skulls, Mesozoic crocodile coprolites, and the skull of the earliest proposed hominid, *Sahelanthropus tchadensis*. Such large fossils present new challenges compared with the tiny rodent teeth or insects in amber. And then there are the unexpected surprises. "Le Moustier has crashed," announced Tafforeau as imaging of Hoffmann's specimen began. The fossil shifted just 5 micrometers on its pedestal of wax, and the plaster used to restore the fossil absorbed too much of the beam, making phase-contrast imaging difficult.

As the week progressed, Tafforeau, Smith, and colleagues worked around the clock to use every minute of beam time. They had better luck with a jawbone of a Neandertal from Krapina, Croatia, which produced sharp images that can be contrasted with those of an early modern human from Qafzeh Cave in Israel. But the answer to their question—how fast these Neandertals grew up—won't be known until after the team has analyzed many terabytes of data.

Paleoanthropologist Jean-Jacques Hublin of the Max Planck Institute for Evolutionary Anthropology watched the images flashing up on a bank of computer screens in the control room and reflected on the march of technology during his lifetime. "When I started my career in paleoanthropology, we used only calipers and a camera," he said. "I never imagined then that we shall time the development of a Neandertal with an accelerator."

–ANN GIBBONS

CREDIT: TANYA SMITH/MPI-EVA

CORRECTED 21 DECEMBER 2007; SEE LAST PAGE

LETTERS

edited by Jennifer Sills

Virtual Worlds, Real Healing

IN THE SOCIAL COGNITION SPECIAL SECTION, G. MILLER EXPLAINED how virtual worlds such as Second Life have become a fertile ground for psychologists exploring human behavior ("The promise of parallel universes," 7 September, p. 1341). In addition to the important social applications mentioned in Miller's article, online communities are playing an emerging role in health services.

Compared with the traditional telehealth systems (i.e., chat, e-mail, and videoconference), online virtual worlds provide the remote user, or patient, with a feeling of embodiment that has the potential to facilitate the clinical communication process and to positively influence group cohesiveness in group-based therapies. It may also create higher levels of interpersonal trust (*1*), which is a fundamental requirement for establishing a successful therapeutic alliance.

Recent evidence has shown that virtual reality–based treatments effectively combat anxiety disorders (*2*) and allow subjects to develop real-world skills starting from virtual experiences (*3*). These successes raise the possibility of creating online immersive therapeutic environments for specific disorders. Imagine, for example, a patient with a social phobia who avoids any interaction with other people. After a number of face-to-face sessions with a therapist, the patient can use his personal avatar to explore a virtual environment, such as a virtual pub in which he can ask the barman for a drink. In the following sessions, other people progressively enter the same virtual pub (they can be other patients, for example) and interact with the patient until he can develop efficient social contacts. The therapist can remotely monitor the patient's psychological, physiological, and emotional responses with the use of biomonitoring systems and can modify the intervention on the basis of the therapeutic needs. This is just one example of the promise of virtual worlds in clinical settings.

Virtual therapy. An example of a group support therapy scenario in Second Life.

ALESSANDRA GORINI

Applied Technology for Neuro-Psychology Laboratory, Istituto Auxologico Italiano, 20146 Milan, Italy.

References

1. G. Bente, S. Rüggenberg, N. C. Krämer, paper presented at the 8th International Workshop on Presence, London, 21 to 23 September 2005.
2. L. Gregg, N. Tarrier, *Soc. Psychiatry Psychiatr. Epidemiol*. **42**, 343 (2007).
3. G. Riva, *CyberPsychol. Behav*. **8**, 220 (2005).

The Age-Old Question of Researcher Innovation

UNFORTUNATELY, BOTH Y. BHATTACHARJEE ("The young and the innovative," ScienceScope, 21 September, p. 1663) and Jeremy Berg, director of the National Institute of General Medical Sciences, perpetuate the myth that "[e]arly-career types are historically the ones who come up with the most innovative ideas." Though this myth remains popular, the available empirical evidence suggests that middle-aged scientists are (i) more apt than young scientists to make revolutionary discoveries (*1*, *2*) and (ii) more productive than young scientists (*3*). In fact, young scientists are not even especially prone to accept a new theory before older scientists (*4*, *5*). It is distressing that funding agencies are making important decisions on the basis of a popular myth that has been examined empirically.

K. BRAD WRAY

Department of Philosophy, State University of New York, Oswego, NY 13126, USA.

References

1. K. B. Wray, *Soc. Stud. Sci*. **33**, 1 (2003).
2. H. Zuckerman, *Scientific Elite: Nobel Laureates in the United States* (Transaction Publishers, New Brunswick, NJ, 1996).
3. S. Cole, *Am. J. Soc*. **84**, 4 (1979).
4. D. L. Hull, P. D. Tessner, A. M. Diamond, *Science* **202**, 717 (1978).
5. P. Messeri, *Soc. Stud. Sci*. **18**, 1 (1988).

Response

IN HIS LETTER, WRAY POINTS OUT THAT EMPIRICAL data suggests that many revolutionary discoveries are made by "middle-aged scientists." He cites studies that reveal that 24 scientists discussed in Thomas Kuhn's "The Structure of Scientific Revolutions" had a mean age of 37.4 when they made their revolutionary contributions (*1*) and that a large collection of Nobel laureates had a mean age of 38.7 at the time of their prize-winning discoveries (*2*). The mean age of the NIH Director's New Innovator Awardees who have just received their first substantial independent research funding from the NIH is approximately 37, somewhat younger than the mean age for new NIH R01 grantees of 41. One of the

motivations for the NIH Director's New Innovator Award (http://grants.nih.govgrants/new_ investigators/innovator_award/), as well as other NIH programs such as the Pathway to Independence Award (http://grants.nih.gov/grants/new_investigators/pathway_independence.htm), is to provide new opportunities for outstanding scientists to launch their independent careers. The empirical data that Wray cites and these NIH grants relate to scientists in the same age range. Whether these scientists are young or middle-aged may be in the eye of the beholder.

JEREMY M. BERG

National Institute of General Medical Sciences, National Institutes of Health, Bethesda, MD 20892–6200, USA.

References

1. K. B. Wray, *Soc. Stud. Sci.* **33**, 1 (2003).
2. H. Zuckerman, *Scientific Elite: Nobel Laureates in the United States* (Transaction Publishers, New Brunswick, NJ, 1996).

Misreading Dr. Venter's Genome

ACCORDING TO THE RECENT NEWS OF THE WEEK article "Venter's genome sheds new light on human variation" by J. Cohen (7 September, p. 1311), the annotation of J. Craig Venter's published genome sequence reveals that Dr. Venter is at increased risk for "antisocial behavior." The gene variant the article is apparently referencing is that of the gene for monoamine oxidase A (*MAOA*) (*1*). The table of variants lists Venter's *MAOA* gene as containing four copies of the uVNTR repeat. Caspi *et al*. (*2*) originally reported that children who were subjected to severe child abuse and who carried the four–uVNTR repeat allele of *MAOA* were less likely to exhibit antisocial behavior than those with three repeats. If these findings are correct, then Venter is at lower risk, not increased risk, for "antisocial behavior."

The confusion is understandable. Attempts to replicate the original Caspi study have yielded mixed results. Despite contradictory findings, these results have been consistent in several studies: (i) Childhood maltreatment is the strongest predictor of antisocial behavior among the subjects of these studies, and (ii) variation in the *MAOA* gene is not predictive of antisocial behaviors later in life (*3*). Perhaps Cohen found it amusing that Venter's genome supposedly predicted risk for antisocial behavior. However, the misinformation spread by such mistakes, or premature conclusions regarding behavioral genetics research, could result in much more damaging consequences for other individuals whose genomes are sequenced, especially if they are not in as secure a position as Dr. Venter.

JON BECKWITH[1] AND COREY MORRIS[2]

[1]Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA. E-mail: jbeckwith@hms.harvard.edu. [2]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA. E-mail: cmorris@hms.harvard.edu

References

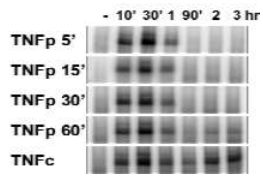
1. S. Levy *et al.*, *PLoS* **5**, 10 (2007).
2. A. Caspi *et al*., *Science* **297**, 851 (2002).
3. C. Morris *et al.*, *GeneWatch* **20**, 2 (2007).

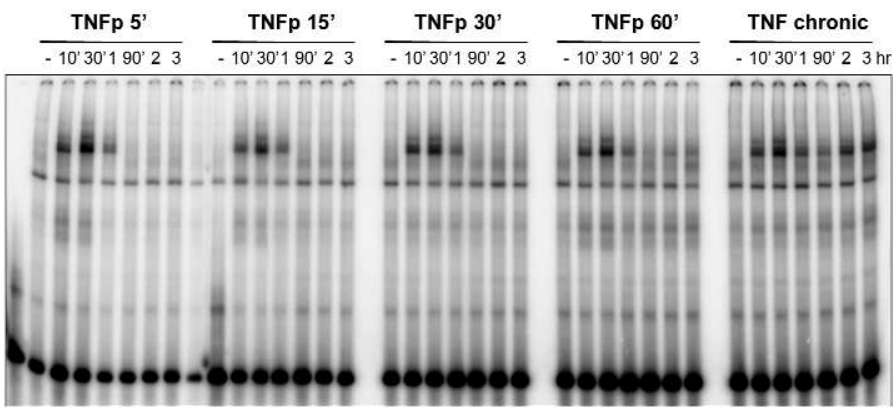
Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

CORRECTIONS AND CLARIFICATIONS

Reports: "The IκB–NF-κB signaling module: Temporal control and selective gene activation" by A. Hoffmann *et al.* (8 November 2002, p. 1241). It has come to our attention that Fig. 3B gives the appearance that lanes might have been spliced or possibly duplicated. The experiments that yielded this figure were carried out in 1997 using autoradiography when the authors were at the Massachusetts Institute of Technology. Similar experiments were rerun after the authors had moved to the California Institute of Technology. Because more stringent standards for handling electronic images have arisen more recently [see, e.g., M. Rossner, K. M. Yamada, *J. Cell Biol*. **166**, 11 (2004)], we provide a recently created figure based on data from a similar experiment (right), as well as an image of the full gel (below) captured with a Molecular Dynamics Phosphoimager.





TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Synthesis of Ultra-Incompressible Superhard Rhenium Diboride at Ambient Pressure"

Natalia Dubrovinskaia, Leonid Dubrovinsky, Vladimir L. Solozhenko

Chung *et al*. (Reports, 20 April 2007, p. 436) reported the synthesis of superhard rhenium diboride (ReB_2) at ambient pressure. We show that ReB_2, first synthesized at ambient pressure 45 years ago, is not a superhard material. Together with the high cost of Re, this makes the prospect for large-scale industrial applications of ReB_2 doubtful.

Full text at www.sciencemag.org/cgi/content/full/318/5856/1550c

RESPONSE TO COMMENT ON "Synthesis of Ultra-Incompressible Superhard Rhenium Diboride at Ambient Pressure"

Hsiu-Ying Chung, Michelle B. Weinberger, Jonathan B. Levine, Robert W. Cumberland, Abby Kavner, Jenn-Ming Yang, Sarah H. Tolbert, Richard B. Kaner

Dubrovinskaia *et al*. question our demonstration that rhenium diboride (ReB_2) is hard enough to scratch diamond. Here, we provide conclusive evidence of a scratch through atomic force microscopy depth profiling and elemental mapping. With high hardness, high-bulk modulus, and the ability to withstand extreme differential stress, ReB_2 and related materials should be investigated regardless of their cost, which is not prohibitive.

Full text at www.sciencemag.org/cgi/content/full/318/5856/1550d

ERRATUM

Post date 21 December 2007

Letters: "Virtual worlds, real healing" by A. Gorini *et al*. (7 December 2007, p. 1549). Two author names were omitted. The complete list of authors is Alessandra Gorini,[1,2] Andrea Gaggioli,[1,3] Giuseppe Riva,[1,3] and their affiliations are as follows: [1]Applied Technology for Neuro-Psychology Laboratory, Istituto Auxologico Italiano, 20100 Milan, Italy. [2]Research Institute Brain and Behaviour, Maastricht University, Netherlands. [3]Psychology Department, Catholic University of Milan, Italy. The authors and affiliations have been corrected in the HTML version on the *Science* Web site.

Comment on "Synthesis of Ultra-Incompressible Superhard Rhenium Diboride at Ambient Pressure"

Natalia Dubrovinskaia,[1,2]* Leonid Dubrovinsky,[3] Vladimir L. Solozhenko[4]

Chung *et al*. (Reports, 20 April 2007, p. 436) reported the synthesis of superhard rhenium diboride (ReB_2) at ambient pressure. We show that ReB_2, first synthesized at ambient pressure 45 years ago, is not a superhard material. Together with the high cost of Re, this makes the prospect for large-scale industrial applications of ReB_2 doubtful.

Research on superhard materials (those with hardness higher than 40 GPa) is driven by both scientific and practical objectives: the desire to understand their structure and bonding, which determine unique properties of these materials, and the demand of modern technologies for robust materials with superior properties. Chung *et al.* (*1*) recently reported the synthesis of superhard rhenium diboride (ReB_2) at ambient pressure. They reported a very high hardness value of 48 GPa and suggested impressive future applications and competitiveness of this material in a row of superhard materials. We question the validity of these claims.

Synthesis of pure ReB_2 by reaction of rhenium with amorphous boron at high temperatures (1200 to 1500°C) and ambient pressure and its crystal structure were reported 45 years ago (*2*). It is not a novel material, and its previous development and discovery was not discussed in the Chung *et al*. article with sufficient emphasis, especially given that one of the three synthesis methods they reported is extremely similar to that used by LaPlaca and Post (*2*).

The ReB_2 Vickers hardness (H_V) of 48.0 (±5.6) GPa was estimated not in the asymptotic-hardness region [Fig. 1; reprinted from figure 2 in (*1*)], as recommended for hard and superhard materials (*3*), but at a very small load that is inappropriate for this class of solids. For soft materials in hardness testing, plastic deformation can be assumed, and the results can be easily interpreted. But for superhard materials, indentation is no longer controlled by plastic deformation alone, and issues such as brittle cracking and deformation of the indenting tip come into play. These effects change the hardness of a material with load, and attempting to infer the hardness of a material above the asymptotic leveling is not informative (*3*). This problem was discussed in detail at the International Workshop on Advanced Superhard Materials (Villetaneuse, France, 10 to 12 December 2003); the recommendations were published as a letter to the scientific community (*3*). For comparison, in our recent study of a superhard boron nitride nanocomposite, hardness reached 145 GPa at low loads, but we reported the asymptotic-hardness value of 83 GPa (*4*) (Fig. 2).

As seen in Fig. 1, the hardness of ReB_2 in the asymptotic-hardness region reaches only 30.1 (±1.3) GPa (*1*), so this phase cannot be considered a superhard one. There are many carbides, nitrides, and borides with similar Vickers hardness [WC, 26 to 28 GPa (*5*); SiC, 27 to 31 GPa (*6*); TiB_2, 33 GPa (*6*); ZrB_2, 35 GPa (*6*)].

Demonstration of the ability of ReB_2 to scratch a diamond surface is also problematic. The optical microscopy image presented by Chung *et al*. (Fig. 1B) gives the impression that this was not a true scratch, but rather a smearing of ReB_2 on the surface of the diamond crystal. If the authors wanted to prove that the ReB_2 indeed scratched the diamond, they should have provided more robust evidence such as an AFM (atomic force microscopy) map of the scratched area. However, even a proven scratch itself does not confirm superior hardness of ReB_2, because it is well known that materials much softer than diamond can damage its surface (*7*). A scratch test is more of a quick field test for identifying minerals and cannot be considered a reliable scientific method in general.

The claim about prospective applications and competitiveness of ReB_2 is questionable from the point of view of both the functional properties of ReB_2 and its commercial value. First, in hardness ReB_2 ($H_V \approx 30$ GPa) cannot even compete with commercially available polycrystalline cBN ($H_V > 40$ GPa), which is successfully used for machining ferrous steels instead of diamond. Second, the cost of raw materials, particularly rhenium [which is six times as expensive as platinum and nine times as expensive as gold (*8*)], is much higher than that of other precursors for superhard materials synthesis. Thus, the prospect of producing hard-tool inserts from ReB_2 seems unrealistic.

Methods of producing superhard coatings at ambient or very low pressure on an industrial scale, including chemical and physical vapor deposition, are well known. The search for alternatives to high pressure–high temper-

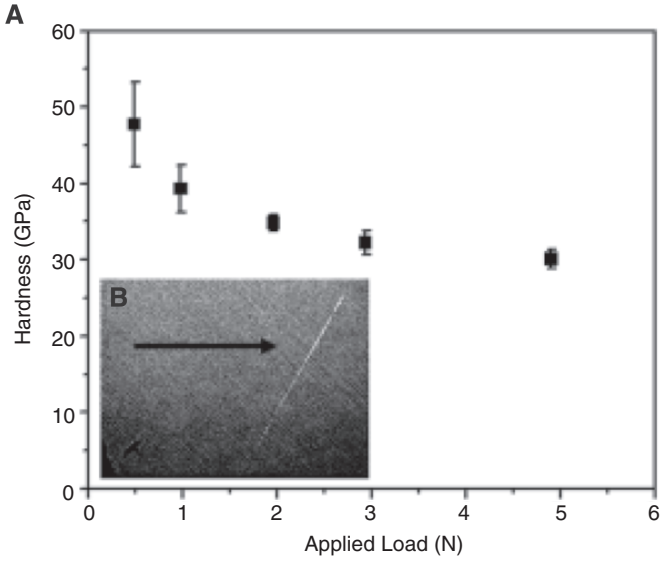


Fig. 1. (**A**) "H_v of ReB2 plotted as a function of applied load measured at room temperature, using a four-sided pyramidal diamond indenter tip. The average hardness increases from 30 to 48 GPa as the applied load decreases from 4.9 to 0.49 N (error bars, ±1 SD). (**B**) An ingot of ReB_2 creates a scratch on the surface of a natural diamond parallel to the (100) plane." [reprinted from figure 2 in (*1*)].

[1]Mineralphysik und Strukturforschung, Mineralogisches Institut, Universität Heidelberg, 69120 Heidelberg, Germany. [2]Lehrstuhl für Kristallographie, Physikalisches Institut, Universität Bayreuth, 95440 Bayreuth, Germany. [3]Bayerisches Geoinstitut, Universität Bayreuth, 95440 Bayreuth, Germany. [4]Laboratoire des Propriétés Mécaniques et Thermodynamique des Matériaux, Centre National de la Recherche Scientifique, Université Paris Nord, 93430 Villetaneuse, France.

*To whom correspondence should be addressed. E-mail natalia.dubrovinskaia@min.uni-heidelberg.de

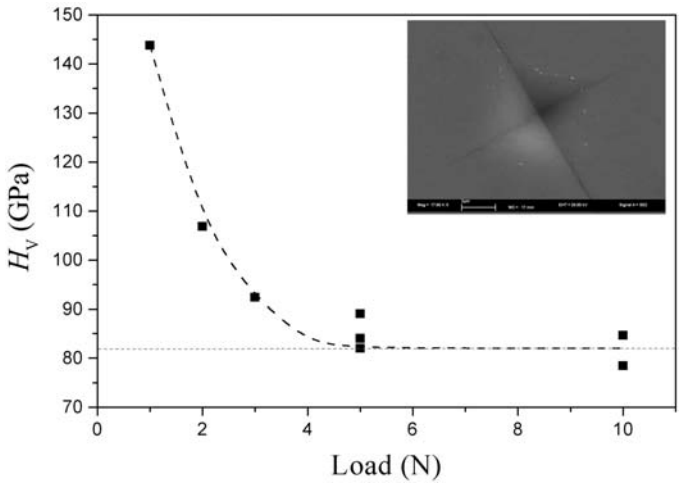


Fig. 2. The load dependence of the Vickers hardness for the sample of superhard BN nanocomposite synthesized at 20 GPa and 1870 K. The inset shows a scanning electron microscopy image of a typical indentation produced by 20-s loading.

ature methods of bulk hard material synthesis continues to be a worthy task.

References

1. H.-Y. Chung *et al.*, *Science* **316**, 436 (2007).
2. S. J. la Placa, B. Post, *Acta Crystallogr.* **15**, 97 (1962).
3. V. Brazhkin *et al.*, *Nat. Mater.* **3**, 576 (2004).
4. N. Dubrovinskaia *et al.*, *Appl. Phys. Lett.* **90**, 101912 (2007).
5. N. P. Bansal, Ed. *Handbook of Ceramic Composites* (Kluwer, Dordrecht, Boston, London, 2005).
6. Y. G. Gogotsi, R. A. Andrievski, Eds., *Materials Sciences of Carbides, Nitrides, and Borides, NATO ASI Series 3, High Technology* (Kluwer, Dordrecht, Boston, 1999).
7. R. Berman, Ed. *Physical Properties of Diamonds* (Clarendon Press, Oxford, 1965).
8. www.taxfreegold.co.uk/preciousmetalpricesusdollars.html

11 July 2007; accepted 29 October 2007
10.1126/science.1147650

Response to Comment on "Synthesis of Ultra-Incompressible Superhard Rhenium Diboride at Ambient Pressure"

Hsiu-Ying Chung,[1,2] Michelle B. Weinberger,[1] Jonathan B. Levine,[1] Robert W. Cumberland,[1] Abby Kavner,[3] Jenn-Ming Yang,[2] Sarah H. Tolbert,[1]* Richard B. Kaner[1,2]*

Dubrovinskaia *et al*. question our demonstration that rhenium diboride (ReB_2) is hard enough to scratch diamond. Here, we provide conclusive evidence of a scratch through atomic force microscopy depth profiling and elemental mapping. With high hardness, high-bulk modulus, and the ability to withstand extreme differential stress, ReB_2 and related materials should be investigated regardless of their cost, which is not prohibitive.

Dubrovinskaia *et al*. (*1*) raise a number of issues regarding our report (*2*) on rhenium diboride (ReB_2) that deserve additional attention. First, we would like to emphasize that we never claimed to be the first group to synthesize ReB_2. That honor indeed goes to La Placa and Post, to whom we gave credit in reference 16 in (*2*). However, we realized, through hardness, incompressibility, and differential stress experiments, that ReB_2 has scientifically interesting mechanical properties.

Dubrovinskaia *et al.* (*1*) express skepticism over the ability of ReB_2 to scratch diamond. They argue that the diamond scratch shown in (*2*) was actually ReB_2 deposited on the diamond surface and that proof of a real scratch would require evidence such as an atomic force microscopy (AFM) image. Here, we provide such proof. An ingot of ReB_2 ~4 mm in diameter was attached to a stylus with mounting wax. The sample was moved across a polished diamond surface using just the weight of the stylus to supply the force. Fig. 1 shows an AFM image of the resulting scratch. The depth profile indicates that the scratch is 2 μm wide, with a depth of ~230 nm. Energy dispersive x-ray (EDX) spectroscopic mapping (Fig. 1, inset) indicates that there is no detectable rhenium deposited on the surface of the diamond. We hope that this new data will end the debate as to whether ReB_2 can scratch diamond. We would further like to point out that a scratch test is not a quantitative method for determining hardness but rather a qualitative test indicating that ReB_2 has mechanical properties worthy of serious investigation.

Dubrovinskaia *et al.* (*1*) also downplay the importance of scratching diamond and make the somewhat misleading statement that "materials much softer than diamond can damage its surface." Although the statement is true, the experiments cited by Berman [reference 7 in (*1*)] result in damage to diamond by either (i) graphitization from the heat induced by a metal ball bearing rotating at speeds in excess of 100 m/s or (ii) formation of radial cracks from tungsten carbide balls applied with loads exceeding 30 N (*3*–*5*). Mechanically, these scenarios are both very different from the deliberate formation of a linear scratch on a surface. To the best of our knowledge, only four bulk materials have previously been reported to scratch diamond, all of which are regarded as superhard: cubic boron nitride, B_6O, fullerite, and diamond-like materials (*6*–*9*).

The comments made by Dubrovinskaia *et al.* do, however, raise the important issue of what it means to be superhard. At low loads, the hardness of many materials (including ReB_2) exhibit a strong dependence on load, increasing as the load decreases. This is known as the indentation size effect. For this reason, Dubrovinskaia *et al.* believe that hardness values calculated in this

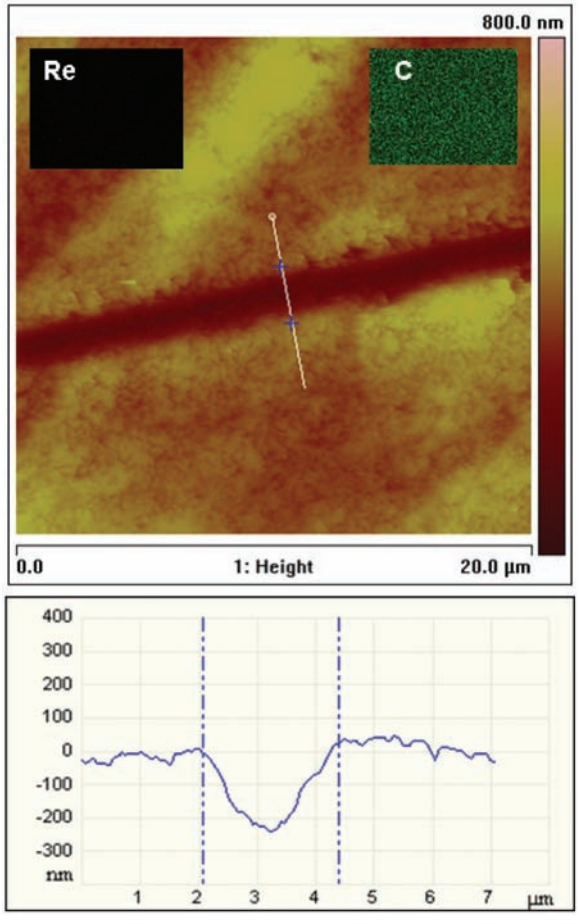


Fig. 1. An AFM image (**top**) and the corresponding depth profile (**bottom**) of a scratch made by ReB_2 on the surface of diamond. The white line follows the depth profile; the blue + s correspond to the dashed lines in the lower image. The scratch has an approximate depth of 230 nm. The insets show elemental density maps for carbon and rhenium over the entire area of the image. The green pixels in the top right inset indicate that carbon is uniformly distributed across the diamond. The absence of Re along the scratch, which should appear as red pixels in the top left inset, clearly demonstrates that ReB_2 has not been deposited on the diamond surface.

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095–1569, USA. [2]Department of Materials Science and Engineering, University of California, Los Angeles, CA 90095–1567, USA. [3]Department of Earth and Space Sciences, University of California, Los Angeles, CA 90095–1567, USA.

*To whom correspondence should be addressed. E-mail: tolbert@chem.ucla.edu (S.H.T.); kaner@chem.ucla.edu (R.B.K.)

regime are meaningless. The asymptotic hardness of ReB_2, with a value of 30.1 GPa, lies well below the generally accepted value of 40 GPa for superhard materials. However, other materials (e.g., transition metal borides and carbides) that have a comparable hardness to ReB_2 in the asymptotic region have not been reported to scratch diamond. Perhaps the low-load data, which achieves its maximum average hardness of 48.0 GPa at 0.49 N, in addition to the anisotropic nature of ReB_2, provide an explanation for its ability to scratch diamond. The one fact that seems clear is that until the indentation size effect is more thoroughly understood, hardness data should be collected as a function of load, and the full load dependence should be reported.

This issue leads to the more general question of how the search for superhard materials should proceed. From our work, it is clear that more than just diamond and diamond-like materials containing first row elements should be considered. Although it is not our specific priority to determine the feasibility or cost-effectiveness of a material for industrial applications, we would like to point out that Dubrovinskaia *et al.* incorrectly report the price of rhenium. At the time of this publication, rhenium metal could be purchased for approximately half the price of gold (*10*). It is clear that substituting other, less expensive transition metals for rhenium is an area that warrants future study.

Having provided clear evidence in support of our previous claims, it should be noted that Dubrovinskaia *et al*. (*11*) have demonstrated a truly remarkable method to increase the hardness of cubic boron nitride by making a nanocomposite. Because this method should be applicable to many other materials, we are now working to synthesize nanocomposites of ReB_2 in hopes of substantially increasing its hardness.

References and Notes

1. N. Dubrovinskaia, L. Dubrovinsky, V. L. Solozhenko, *Science* **318**, 1550 (2007); www.sciencemag.org/cgi/content/full/318/5856/1550c.
2. H.-Y. Chung *et al*., *Science* **316**, 436 (2007).
3. R. Berman, Ed., *Physical Properties of Diamond* (Clarendon Press, Oxford, 1965).
4. F. P. Bowden, E. H. Freitag, *Proc. R. Soc. Lond. A* **248**, 350 (1958).
5. S. Tolansky, V. R. Howes, *Proc. Phys. Soc.* **70**, 521 (1956).
6. R. H. Wentorf Jr., *J. Chem. Phys.* **26**, 956 (1957).
7. A. R. Badzian, *Appl. Phys. Lett.* **53**, 2495 (1988).
8. V. Blank *et al*., *Diamond Relat. Mater.* **7**, 427 (1998).
9. N. Dubrovinskaia *et al*., *Appl. Phys. Lett.* **87**, 83106 (2005).
10. Dubrovinskaia *et al.* claim that rhenium costs nine times as much as gold, citing a Web site that simply attempts to estimate the free market price of the metal, which has no bearing on the current cost of rhenium. Rhenium metal can be purchased from Rhenium Alloys, Inc. at a price of $12/gram.
11. V. Dubrovinskaia *et al*., *Appl. Phys. Lett.* **90**, 101912 (2007).

31 July 2007; accepted 6 November 2007
10.1126/science.1147704

FOR YOUNGER READERS

Science Books for Fun and Learning—Some Recommendations from 2007

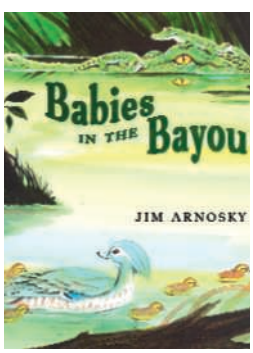
Are you looking for holiday gifts for children or young adults whose interest in science you are trying to encourage? We offer as suggestions the finalists for the 2008 *Science Books and Films* Prizes for Excellence in Science Books. The prizes honor books that promote an understanding and appreciation of science in younger readers. Sponsored by the AAAS and Subaru, they are awarded in four categories: children's science picture book (for readers in grades K–4), middle grades science book (grades 5–8), young adult science book (high school), and hands-on science/activity book (any age). This year none of the four finalists for the young adult award was specifically intended for that age group—all were written for the general public. The titles considered for the 2008 prizes were published between September 2006 and August 2007.

Here, we present our short descriptions of the 17 finalists chosen by panels of librarians, educators, and scientists. Full reviews of each book have been published or will appear in *Science Books and Films*, and AAAS members can read these reviews on the Web. The four winners for 2008 will be announced at the AAAS Annual Meeting in Boston in February.

The criteria for evaluating the books include a clear and accurate presentation of scientific concepts. But we join the judges in hoping that the finalists will entice young readers to turn to science books for enjoyment as well as for information.

–Heather Malcomson,[1] Sherman Suter, and Barbara Jasny

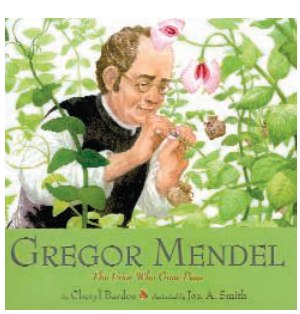
Children's Science Picture Book



Babies in the Bayou. Jim Arnosky. Putnam (Penguin), New York, 2007. 32 pp. $16.99, C$21. ISBN 9780399226533.

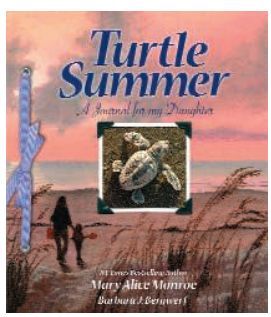
The winner of a 2005 *Science Books & Films* prize for his lifetime contributions to illustration of children's science books, Arnosky highlights his gift for bringing the natural world to life in yet another outstanding book. His beautiful illustrations and simple text introduce the youngest of readers to the important predator-prey relationship. The charming and accurate representations of bayou animals will help foster an appreciation of the natural world in any child.

Gregor Mendel. The Friar Who Grew Peas. Cheryl Bardoe, illustrated by Jos. A. Smith. Abrams Books for Young Readers (Abrams), New York, in association with the Field Museum, Chicago, 2006. 34 pp. $18.95, C$26.95, £10.95. ISBN 9780810954755.



Bardoe uses pictures and words to describe the life and work of Mendel. She graphically portrays his willingness when in his youth to make sacrifices in order to study. Her particularly effective account of his research on peas should paint a clear picture of heredity for young readers.

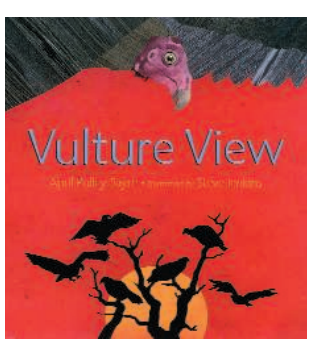
Turtle Summer. A Journal for My Daughter. Mary Alice Monroe and Barbara J. Bergwerf. Sylvan Dell, Mt. Pleasant, SC, 2007. 32 pp. $15.95. ISBN 9780977742356. Paper, $8.95. ISBN 9780977742370.



Late each spring, female loggerhead turtles return to lay their eggs in sandy beaches along the coast of South Carolina. Monroe introduces young readers to these endangered sea turtles, other seashore fauna, and the efforts of volunteers to watch and care for the nests and thus increase the numbers of hatchlings that set off midsummer for decades of growth at sea. Her "scrapbook" is filled with Bergwerf's snapshots of nesting activity, young turtles, shells, and birds. It effectively invites children to observe and interact with nature.

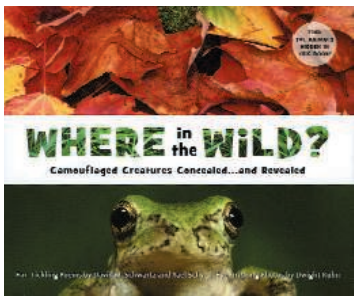
Vulture View. April Pulley Sayre, illustrated by Steve Jenkins. Holt, New York, 2007. 32 pp. $16.95, C$21. ISBN 9780805075571.

The eating habits of vultures are portrayed by Sayre in a way that is sure to delight young children who love gross and squishy things. Using a very appealing rhyming format, she follows the birds as they soar and scan, searching for food: "Vultures like a mess. / They land and dine. / Rotten is fine." The illustrations also bring out other parts of the vulture lifestyle, such as bathing and roosting. It is good to see elegant artwork featuring a species that is conventionally considered unattractive. A section at the back of the book provides additional information about vulture biology.



CREDIT: PETER HOEY

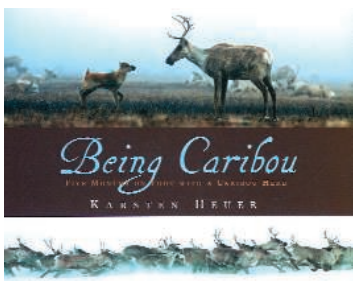
[1] *Science Books and Films*, 1200 New York Avenue, NW, Washington, DC 20005, USA. E-mail: hmalcoms@aaas.org



Where in the Wild? Camouflaged Creatures Concealed and Revealed. David M. Schwartz and Yael Schy, photographs by Dwight Kuhn. Tricycle (Ten Speed), Berkeley, CA, 2007. 24 pp. $15.95. ISBN 9781582462073.

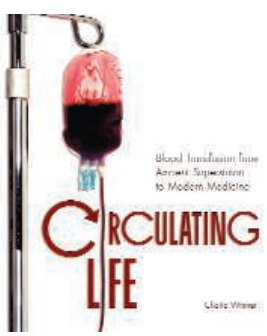
One must look carefully to spot the creatures that appear in this book. Each has colors and patterns that allow it to blend into the surroundings of its natural habitat. Ten short verses by Schwartz and Schy hint at the identity or location of the species hidden in Kuhn's photographs. If one gets frustrated by an animal's skillful deception, unfolding the folio page reveals its position in a faded version of the photo. Accompanying text introduces the animal's natural history and explains how it uses color and behavior to help escape being eaten or capture food.

Middle Grades Science Book



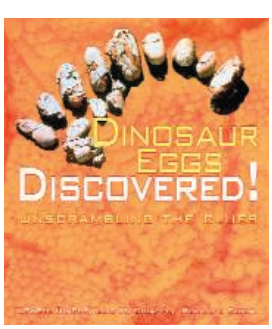
Being Caribou. Five Months on Foot with a Caribou Herd. Karsten Heuer. Walker, New York, 2007. 50 pp. $17.95. ISBN 9780802795656.

Full of real-life adventures, this book recounts the experiences of wildlife biologist Heuer and his wife after they joined more than 100,000 caribou on a 1000-mile trek to Alaska's Arctic National Wildlife Refuge. The couple endures brutal weather and physical hardships to study and document the herd's amazing journey across the northernmost reaches of North America. Heuer describes caribou movements, feeding, birthing, playing, and predator avoidance. At the end of the book, readers will feel exhausted by the expedition and concerned about the plight of caribou.



Circulating Life. Blood Transfusion from Ancient Superstition to Modern Medicine. Cherie Winner. Twenty-First Century (Lerner), Minneapolis, MN, 2007. 112 pp. $30.60. ISBN 9780822566069. Discovery!

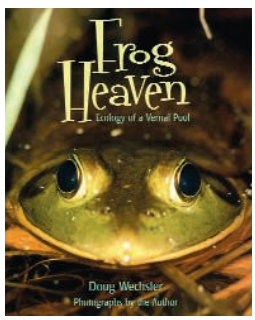
Providing a thorough history of blood transfusion, Winner starts with the earliest ideas about blood, progresses through the first unsafe and unenlightened attempts, and ends with current practices and a look at the prospects for synthetic blood. Along the way, readers will learn a great deal about blood itself and about issues of blood safety. The glossary, bibliography, and list of Internet resources will aid those seeking more information.



Dinosaur Eggs Discovered! Unscrambling the Clues. Lowell Dingus, Luis M. Chiappe, and Rodolfo Coria. Twenty-First Century (Lerner), Minneapolis, MN, 2008. 112 pp. $30.60. ISBN 9780822567912. Discovery!

This excellent book by three vertebrate paleontologists starts with an amazing find—a field of fossilized dinosaur eggs in Argentina. The authors proceed to explain how they were able to answer fundamental questions about their discovery, including: "Who laid the eggs?" "When were the eggs laid?" "Were the eggs laid in nests?" "What catastrophe killed the embryos?" Although filled with facts, the book also presents the entertaining story of a scientific expedition.

Frog Heaven. Ecology of a Vernal Pool. Doug Wechsler. Boyds Mills Press, Honesdale, PA, 2006. 48 pp. $17.95. ISBN 9781590782538.

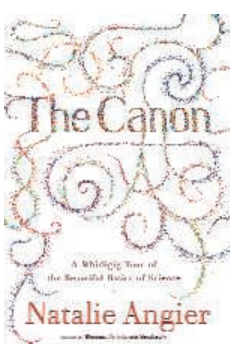


Vernal pools are shallow, seasonal ponds that are not linked to permanent streams and disappear for a while during most years. Thus, they usually lack fish, which makes them an ideal habitat for frogs, salamanders, and insect larvae. Wechsler describes the action at one such pool in a Delaware woods from its filling by autumnal rains, to the spring cacophony of male frogs seeking mates, and on through the inhabitants' race against time as the summer sun dries up the pond. He weaves intriguing details about the biota and his own photos into this informative story of an underappreciated ecosystem.

Tracking Trash. Flotsam, Jetsam, and the Science of Ocean Motion. Loree Griffin Burns. Houghton Mifflin, Boston, 2007. 64 pp. $18, C$24.50. ISBN 9780618581313. Scientists in the Field.

In 1990, thousands of sneakers were washed off a container ship at a spot in the North Pacific. Two years later, thousands of plastic tub toys were similarly spilled at another location south of the Aleutians. Burns explains how Curt Ebbesmeyer and other oceanographers (aided by beachcombers) have used the dispersal of these and other drifting objects to illuminate ocean currents. She also highlights the problems that long-lived plastic trash poses across vast expanses of ocean and on beaches (even very isolated ones) around the world. And she warns of the deadly effects that abandoned ghost nets have on pelagic biota and reefs.

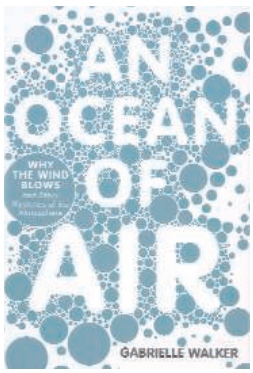
Young Adult Science Book



The Canon. A Whirligig Tour of the Beautiful Basics of Science. Natalie Angier. Houghton Mifflin, Boston, 2007. 303 pp. $27, C$35.95. ISBN 9780618242955. Forthcoming from Faber and Faber, London. £16.99. ISBN 9780571239719.

A Pulitzer Prize–winning science columnist for *The New York Times*, Angier here explores the basic principles of the scientific method and the disciplines of astronomy, biology, chemistry, geology, and physics. Using both anecdotes about and discoveries by scientists, she weaves easy-to-understand explanations of contemporary science. She also gives personal accounts of how various aspects of abstract scientific findings play into everyday life. Seemingly dull concepts such as hydrogen

bonding and exothermic reactions are described in an interesting, metaphor-rich way. Although written for a general audience, the book will appeal to young adults wanting a broader knowledge of important scientific concepts.

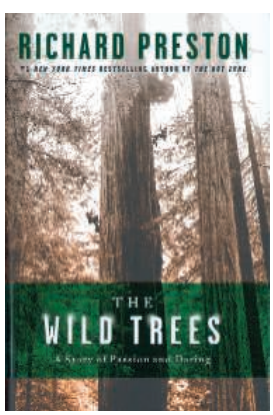


An Ocean of Air. Why the Wind Blows and Other Mysteries of the Atmosphere. Gabrielle Walker. Harcourt, New York, 2007. 288 pp. $25. ISBN 9780151011247. Bloomsbury, London. 321 pp. £15.99. ISBN 9780747581901.

Walker's clear and witty writing makes this popular science book appealing and accessible to high school students. The author is passionate about her subject, referring to the atmosphere as "the single greatest gift our planet possesses." The book is full of interesting profiles of scientists who have spent their careers uncovering the secrets of the atmosphere. Readers will encounter familiar names and stories (e.g., the discoveries of Joseph Priestley and Robert Boyle) but also the work of lesser-known scientists such as William Ferrel and Kristian Birkeland. Walker even includes some nonscientists, like pilot Wiley Post, who first discovered the high, fast-moving winds that flow through Earth's stratosphere.

Is Pluto a Planet? A Historical Journey Through the Solar System. David A. Weintraub. Princeton University Press, Princeton, NJ, 2007. 266 pp. $27.95, £16.95. ISBN 9780691123486.

In ancient times, the Sun and the Moon were among the seven wandering stars, the planets. Weintraub discusses how the concept of planet has changed from then to the present, when astronomers are identifying objects orbiting distant stars. He describes the rises and falls in the number of planets recognized in our solar system—changes that lead him to term Pluto "the fourth ninth planet." (He also notes that Pluto was not the object its discoverer was looking for and that object does not exist.) Although readers may not accept Weintraub's answer to the titular question, they will find his thought-provoking account provides ample information for supporting a variety of positions in the continuing debate.

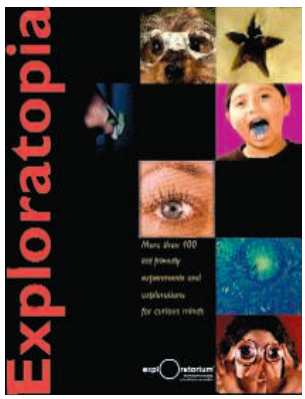


The Wild Trees. A Story of Passion and Daring. Richard Preston. Random House, New York, 2007. 315 pp. $25.95, C$32. ISBN 9781400064892. Allen Lane, London. £20. ISBN 9780241141847.

Novelist Preston, the author of *The Hot Zone*, has turned his attention to California's coastal redwoods, the ecosystems that are maintained in them, and people who were passionately committed to finding and studying the tallest of these trees. The book is meant for a sophisticated young adult (or adult) reader—the protagonists are real, and somewhat eccentric, people who find huckleberries to eat in the canopies of these enormous trees, sleep (and make love) in special hammocks, and sometimes do crazy things that get themselves injured or killed. It will reward anyone who enjoys adventurers, loves climbing, or loves trees.

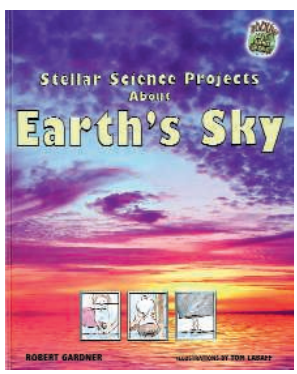
Hands-On Science/Activity Book

Exploratopia. Pat Murphy, Ellen Macaulay, and the Staff of the Exploratorium. Little, Brown, New York, 2006. 383 pp. $29.99, C$36.99. ISBN 9780316612814.



In this natural extension of the San Francisco–based Exploratorium: The Museum of Science, Art, and Human Perception, Murphy, Macaulay, and colleagues of theirs at the museum offer more than 400 kid-friendly experiments and explorations for curious minds. The result is like a tour through the Exploratorium itself. The book consists entirely of hands-on activities, most of which require only easily obtainable everyday materials. It meets all the requirements for a great science activity book: interesting investigations, clear instructions, vivid illustrations, and a bit of humor to hold the attention of kids of all ages.

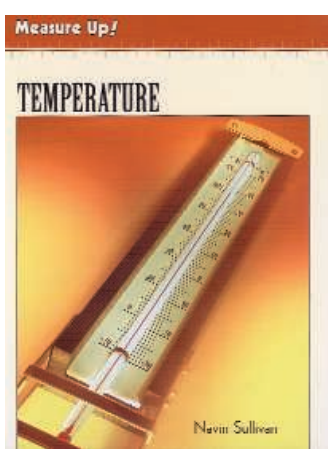
Stellar Science Projects About Earth's Sky. Robert Gardner, illustrations by Tom LaBaff. Enslow Elementary (Enslow), Berkeley Heights, NJ, 2007. 48 pp. $23.93, £11.51. ISBN 9780766027329. Rockin' Earth Science Experiments.



With this basic hands-on book, Gardner aims to help children learn about air pressure, sunlight, and clouds as well as build science fair projects. The book's main strength lies in his clear explanations of the science underlying the observations. I was surprised to find, after guiding two children through innumerable home experiments, a revealing procedure for demonstrating why sunsets are red that I had never encountered before. Some of the experiments will require parental supervision, but the materials are easy to obtain and the further reading includes Internet sites.

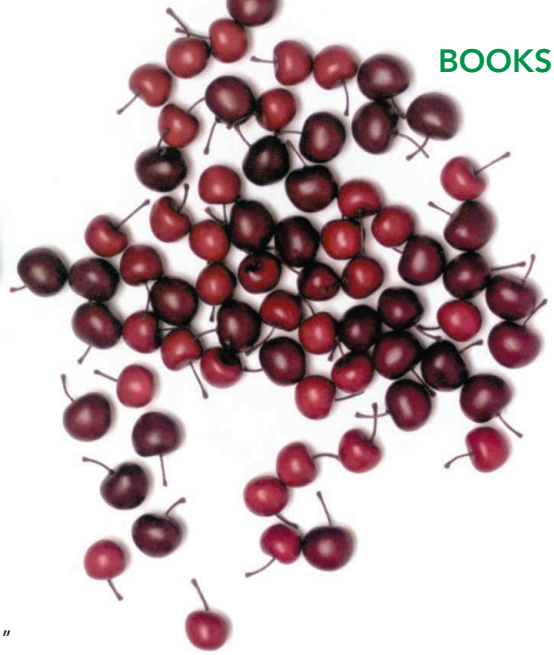
Temperature. Navin Sullivan. Marshall Cavendish Benchmark, New York, 2007. 48 pp. $29.93, £14.59. ISBN 9780761423225. Measure Up!

Sullivan begins by describing heat energy and the ways in which it can travel. After providing a lucid explanation of the difference between the amount of heat in an object and its temperature, he discusses temperature scales, devices for measuring temperature, and temperature effects on people. The book is one in a series for middle-grade readers that explores measurements and their histories. It includes instructions for experimenting with freezing points and for constructing a simple thermoscope.



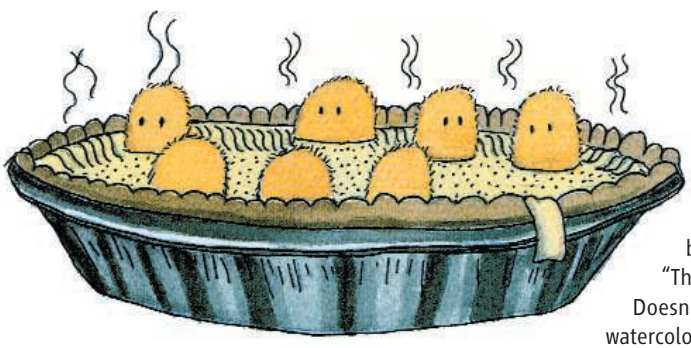
10.1126/science.1153007

BROWSINGS



Great Estimations. Bruce Goldstone. Holt, New York, 2006. 32 pp. $16.95, C$21. ISBN 9780805074468.

Many books for young readers attempt to give them a feeling for small numbers, especially the digits between one and ten. In contrast, Goldstone aims to help children quantify more numerous objects. His first few photographs mix groups of familiar things (such as pencils, die, and paperclips) to show what 10, 100, and 1000 look like. He then presents pairs of known amounts followed by a challenge: for example, 10 cherries, 100 cherries, and (right) "how many cherries are in a quart?" He proceeds to techniques such as clump counting and box and count, and he complicates the pictures by using objects at different distances or of various sizes. In one case, jelly beans in a fish bowl, one must account for unseen items. By training your eyes, the book allows you to obey the author's injunction, "Don't count—estimate!"

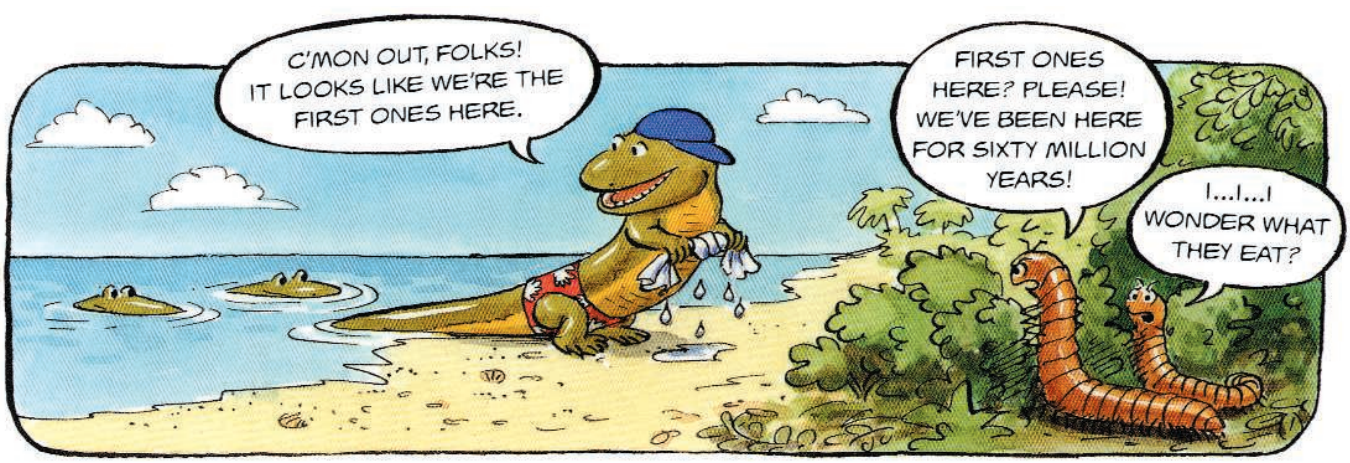


Germ Stories. Arthur Kornberg, illustrations by Adam Alaniz, photography by Roberto Kolter. University Science Books, Sausalito, CA, 2007. 84 pp. $22.50. ISBN 9781891389511.

This collection of humorous yet informative stories comprises rhymes that the late Nobel Laureate Kornberg originally wrote to entertain his grandchildren. Many of the stars of his "germ parade" are harmful pests, like the food-poisoning *Staphylococcus aureus* (left)—of which he writes "In the pie germs grew and frolicked / And spewed out poisons that can cause colic." Others, such as some members of our "intestinal menagerie," he describes as "good guests": "These bugs make vitamins and digest food, / And their great number / Doesn't let bad germs intrude." The poems are illustrated with whimsical watercolors by Alaniz. Photographs, electron micrographs, and a glossary offer additional details for readers intrigued by the microbiota.

When Fish Got Feet, Sharks Got Teeth, and Bugs Began to Swarm. A Cartoon Prehistory of Life Long Before Dinosaurs. Hannah Bonner. National Geographic, Washington, DC, 2007. 41 pp. $16.95, C$21.95. ISBN 9781426300783.

This prequel to *When Bugs Were Big, Plants Were Strange, and Tetrapods Stalked the Earth* spotlights inhabitants and evolutionary events of the Silurian and Devonian periods. Bonner begins about 430 million years ago, when the seas were home to 2-m-long scorpion relatives and small jawless fish. On land, tiny arthropods crept about under short relatives of the mosses. She discusses how plants and animals adapted to land, the birth of dirt, and the first forests. After charting the diversification race among four groups of fish with jaws, she depicts lobe-fin fish going ashore (below) and the skeletal changes that occurred in the transition to tetrapods. For readers who wish to delve deeper into these 80 million years of the middle Paleozoic, she suggests several sites on the Web and provides a comprehensive list of references for her text and reconstructions.



CREDITS (TOP TO BOTTOM): ARNOLD AND DAVID KATZ/BRUCE GOLDSTONE/FROM *GREAT ESTIMATIONS*; ADAM ALANIZ/FROM *GERM STORIES*; HANNAH BONNER/FROM *WHEN FISH GOT FEET*...

ECOLOGY

China's Forestry Reforms

Guangyu Wang,[1] John L. Innes,[1]* Jiafu Lei,[2] Shuanyou Dai,[2] Sara W. Wu[3]

Forestry management policies in China have changed direction to encourage sustainability while balancing land-use, economic growth, and demand for forest products.

Since 2000 (*1*), there have been rapid and extensive changes in forestry policy in China. Investments in the forestry sector since 2000 have exceeded the total investments in the period 1949–99. For its Six Key Forestry Programs (SKFPs) alone, China has invested 183.5 billion renminbi (RMB) (ca. U.S.$22 billion) in the last 6 years and will invest a further 539.8 billion RMB (ca. U.S.$68 billion) in the next 4 years (table S1 in supporting online material). Here, we provide an update on the major forestry reforms introduced since 2000.

Massive investment in the SKFPs, strong demand for wood, and increasing pressure from environmental groups has led to calls for reform of forest ownership. Forests are considered the last battleground for much-needed land-tenure reforms in China, where old laws and practices still present a major barrier to the development of China's forest estate. In 2004, several provinces in the south began to reform forest ownership policies, introducing cuts in forest taxes, free-market mechanisms for forest asset transfers, and private support systems for forestry. These reforms are intended to improve forest infrastructure, enhance the competitive power of Chinese wood products, and improve environmental quality.

China is facing many problems that affect social harmony, including growing pressure on the environment and natural resources. Past government policies have favored economic growth over the environment, but the central government has now proposed a science-based approach to development designed to realize balanced sustainable development (*2*). However, in practice, local governments continue to put economic growth ahead of any concern for the environment, which has led some critics to call for stronger central government control.

China's rapid economic growth, increased capital investment, and growing middle-class consumption have driven up the demand (and prices) for wood products. China not only needs wood to meet domestic demand, it also has a growing and very successful export industry. In 2006, the forest products trade in China was worth U.S.$47.07 billion, a 23% increase over 2005. Forest product imports were valued at $19.39 billion (a 10% increase over 2005) and exports at $27.68 billion (a 34% increase). The trade in the first 6 months of 2007 was valued at $27.2 billion, a 35% increase over the same period in 2006 (*3*). By 2006, China had emerged as the world's largest exporter of furniture, accounting for 43% of U.S. and 33% of European wood furniture imports (*4*). To meet the growing international demand for sustainability assurances in the production of forest products, China is developing a national certification standard and will seek endorsement of its standard by the international Program for the Endorsement of Forest Certification (PEFC).

Forest police patrol in a protected forest area. China has about 60,000 specially trained forest police to enforce policies such as the logging ban.

The Six Key Forestry Programs

The SKFPs cover more than 97% of China's counties and target 76 million hectares of land for afforestation. The Natural Forest Protection Program (NFPP) was introduced in 1998 after a logging ban prompted by the most devastating floods in Chinese history (*1*). After a series of pilot studies, five additional programs were established to promote a more sustainable forest policy (table S2).

Advances and successes. During the past 8 years, the NFPP has brought 98 million ha of forest under effective protection. Logging natural forest has been banned in the upper reach of the Yangtze River and in the middle and upper reaches of the Yellow River. Timber production in the Northeast and Inner Mongolia has been successfully reduced from 18.24 million m^3 in 1997 to 10.99 million m^3 in 2006 (*6*), and 0.67 million displaced forestry workers have been resettled (*5*).

There has also been significant progress in afforestation, with 28 million ha of plantations established in the past 6 years (*6*). The Conversion of Cropland to Forest Program (CCFP)—which pays farmers to plant trees rather than crops—has converted 8.8 million ha of cropland into forests (*6*). Under the CCFP, soil erosion has been reduced by 4.1 million ha, representing a 4.1% annual reduction. For the first time since the establishment of the People's Republic of China, desertification has been reversed, from an annual increase of 3436 km^2 at the end of the 20th century, to the current annual reduction of 1283 km^2 (*5*). This has been largely achieved through the Sand Control Programs for areas in the vicinity of Beijing and Tianjin, the Three-North Shelterbelt Development Program and the Shelterbelt Development Program along the Yangtze River Basin programs. During 2001–06, 831 natural reserves were created, and 19.5 million ha of forestland and special sites were protected under the Wildlife Conservation and Nature Reserves Development Program (*6*).

[1]Faculty of Forestry, University of British Columbia, Vancouver, BC Canada V6T 1Z4. [2]State Forestry Administration, Beijing, China, 100714. [3]World Forest Institute, Portland, OR 97221, USA.

*Author for correspondence. E-mail: john.innes@ubc.ca.

The total area of plantations in China now amounts to 53 million ha, with forest cover increasing from 16.6 to 18.2%, and the forest stock volume increasing from 11.567 billion m^3 to 12.456 billion m^3 since the start of the SKFPs (*7*).

Problems and obstacles. The booming economy has placed greater pressure on a system not yet capable of balancing the growth in wood demand with environmental needs and social justice. Although the central government has been proactive in trying to improve China's forestry basis, the on-the-ground effects at the state and local levels have been mixed (see figure, p. 1556). For example, the central government has been providing major funding for tree-planting, but local governments lack the funding to implement the programs effectively (*8*).

Transfer of responsibilities to local governments means that there is no guarantee of continued funding for the stewardship of the new forests. It is also unclear whether resettled workers and local farmers are directly benefiting from some of the projects. In areas covered by the logging ban, the decline of community services may have exacerbated their economic difficulties. Local corruption is widespread and underregulated corporations have been accused of usurping user rights and failing to compensate farmers for their land.

Reforming China's complex system of forest ownership and user rights is critical to the long-term implementation of its forestry programs. Land ownership reforms will provide farmers with rights to plant trees for income and will give incentives to protect forests. The reforms involve transfer of land to individuals or companies, and compensation packages for those not receiving land. In the CCFP program, the delay in ownership reform has resulted in farmers planting their forest land even though they have no property rights. In some areas impacted by the NFFP program, the needs of local people have been inadequately considered and compensation levels have been too low to offset their losses. Progress in the Forest Industrial Base Development Program, which focuses on fast-growing and high-yielding commercial timber plantations, has been slow, with uncertainty over forestland ownership, resulting in only 0.19 million ha of new plantations established in the last 6 years (*6*).

Ownership Reforms and Auxiliary Policies

Forest ownership reform started in Fujian and Jiangxi provinces and has been extended to the provinces of Zhejiang, Liaoning, Heibei, Shandong, Anhui, and Guangdong.

The central government has removed or reduced forestry taxes to encourage tree planting and forest products manufacturing. Local governments have removed provincial taxes and some fees on forest products. For example, Fujian province has reduced forest product taxes and fees from 46% of the total sale price to 26%. Simultaneously, the government is using transfer payments to support local governance organizations that used to be financed by forest taxes and fees. The Jiangxi provincial government lost $182.5 million in tax revenue but had this sum supplied instead by transfer payments. As a direct result of this change, the average annual cash income for each farmer increased by 13%, or just over $10 (*9*).

To provide a mechanism for the trading of forest assets—land and timber—China established its first pilot futures market, the Fujian Yong'an Forestry Elements Market in 2004. The market consists of a forest and forestland registration center, a forest resource evaluation center, a timber and bamboo exchange, a legal and technical service center, and a labor training center. By May 2007, the market had bought and sold 20,766 ha of forest and provided purchasing loans worth $63.8 million (*10*). In Jiangxi province, there are now 36 such markets established or being set up, and the number of deals has exceeded 3000, valued at $120 million (*9*).

Future Forest Management Structures

The Chinese government is beginning a new phase of forestry reforms intended to open the forest sector to much greater individual and corporate participation, largely through private sector financing. This represents a major break from the past, when most forestry activities were managed through the government. It aims to increase China's forest cover to 26% by 2050, to improve environmental quality, and to develop a competitive forest industry that depends largely on a domestic fiber supply.

To achieve these goals, several changes in policy are being instituted (*11*), beginning with the separation of ecological and commercial forests, each having separate management policies. However, the policies for managing ecological forests and commercial forests are not yet fully in place and need to be integrated with sustainable forest management systems.

The government will strictly protect ecological forests, increasing fire, pest, and biodiversity protection and preventing logging or the conversion of ecological forests to other uses. Local communities and farmers will be compensated if their land is classified as ecological forest. On commercial forests, the government will grant much greater leeway to develop management plans and will allow farmers the freedom to determine harvest age (based on economic maturity), apply intensive forest management, select tree species, pursue economic benefits, and harvest on their own timetable based on agreed forest management plans. The government will also allow regional planners to use private funding to achieve these goals. The government will no longer control, but rather, encourage, the development of the commercial wood products trade.

Although the reforms represent a major shift in policy, the government will continue to be the ultimate authority in regional planning, zoning, and policy direction. The government will still govern forest asset ownership and transference rights, such as issuing licenses for land-use rights, forest ownership, and ownership exchange. It will set regulations to require forest practices to follow sustainable forest management and will encourage the private or public sector to fill gaps to provide services for forest management, such as management consultation, road-building, nurseries, wood markets, and logging.

References and Notes

1. P.-C. Zhang *et al.*, *Science* **288**, 2135 (2000).
2. J. Ma, "A path to environmental harmony," Chinadialogue (30 November 2006); www.chinadialogue.net.
3. Q. Y. Cao, State Forestry Administration Press Conference [in Chinese], www.forestry.gov.cn/xwfbh/xwfbh070912.asp
4. UNECE Timber Committee, *Statement on Forest Products Markets in 2006 and Prospects for 2007* (Report ECE/TIM/06/N01, UNECE Timber Committee, Geneva, 2006).
5. State Forestry Administration, *Enhancing Forestry Ecological Improvement and Accelerating Development of the Industry* (State Forestry Administration, Beijing, 2006).
6. State Forestry Administration, *China Forestry Development Report* (China Forestry Publishing House, Beijing, 2005-07).
7. State Forestry Administration, *The Sixth National Survey on Forestry Resources, Progress Reports* (China Forestry Publishing House, Beijing, 2005).
8. State Forestry Administration, *A Report for Monitoring and Assessment of the Social-Economic Impacts of China's Key Forestry Programs* (China Forestry Publishing House, Beijing, 2003–06).
9. Six Joint-Departmental Investigation Task Force, *Investigation Report on Jiangxi Forest Ownership Reform* (State Forestry Administration, Beijing, 2007).
10. X. X. Sun, "Fujian forest-ownership reform on-the-spot report," *Chinanews*, 6 July 2007, p. 15 [in Chinese]; www.chinanews.com.cn/cj/kong/news/2007/07-06/973437.shtml
11. Central Committee of the Communist Party of China and China State Council, 2003, "Directive to enhance forestry development," issued 25 June 2003; Xinhua News Agency, Beijing, 11 September 2003 [in Chinese], http://news.xinhuanet.com/zhengfu/2003-09/11/content_1075042.htm.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1556/DC1

10.1126/science.1147247

BIOCHEMISTRY

How Some Pili Pull

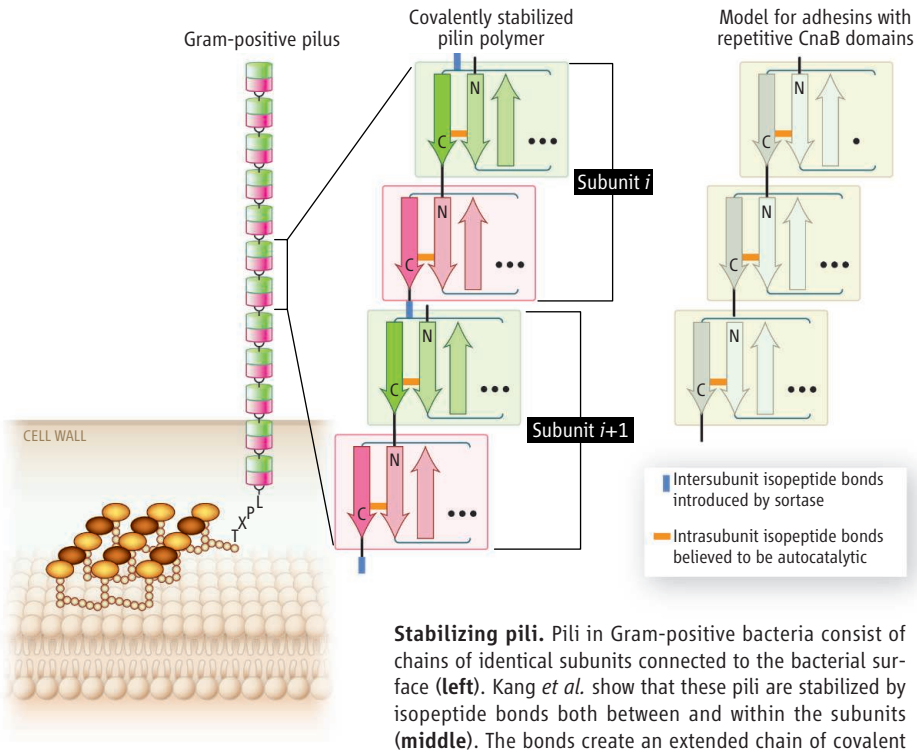
Unexpected covalent bonds explain the high mechanical strength of the ultrathin pili of Gram-positive bacteria.

Todd O. Yeates and Robert T. Clubb

Many bacteria persist in their environmental niches by attaching themselves via hairlike extensions that project from their surfaces. In Gram-negative bacteria, these filamentous, multisubunit protein structures—called pili or fimbria—can also be involved in the transfer of genetic material, induction of signaling in host cells, and twitching motility. Insights into pilus structure and biogenesis in Gram-positive bacteria have begun to emerge only recently (*1*–*3*). Pili in Gram-positive microbes can extend several micrometers from the cell surface but are especially thin, being only a few nanometers thick. On page 1625 of this issue, Kang *et al.* (*4*) show how these thin structures can withstand the mechanical rigors of life outside the cell.

Genomic sequence gazing and several recent experimental studies suggest that pili are widespread across Gram-positive bacteria, including pathogenic *Streptococcus* species (*1*, *2*). Among Gram-positive bacteria, the structure and biogenesis of pili in *Corynebacterium diptheriae* are best understood; they are assembled from three types of pilin subunits, which are joined together to form filaments by a set of cross-linking reactions and are then attached to the cell surface (*5*). The main body of the pilus is formed by a single pilin protein, many copies of which are polymerized into a chain by a transpeptidase enzyme from the sortase family.

In the current model, the sortase cleaves a Leu-Pro-X-Thr-Gly motif near the C-terminal end of one subunit, and then catalyzes the formation of a covalent isopeptide bond between the resultant C-terminal threonine and a conserved lysine side chain that resides in the "pilin motif" of another subunit. The continuation of this process creates a string of polymerized subunits on the cell surface (see the figure, left). The biogenesis of the pilus is completed when the chain of pilin subunits is transferred to a "housekeeping" sortase enzyme, which in turn attaches the chain to the bacterial cell wall (*6*). The isopeptide linkage between pilin subunits is a strategy for pilus stabilization distinct from those used by Gram-negative bacteria, but the structural details were unclear until now.



Stabilizing pili. Pili in Gram-positive bacteria consist of chains of identical subunits connected to the bacterial surface (**left**). Kang *et al.* show that these pili are stabilized by isopeptide bonds both between and within the subunits (**middle**). The bonds create an extended chain of covalent connectivity, thereby strengthening the pilus. Similar bonding features appear to be present in other cell surface adhesion proteins with related three-dimensional folds (**right**).

Kang *et al.* now report the crystal structure of the major pilin subunit from the human pathogen *S. pyogenes*. This pilin protein promotes adhesion to the pharynx and is an attractive target for vaccine development; it is one of the antigens used for more than 50 years to classify *Streptococcus* serotypes (*7*, *8*). Within the crystals, the pilin subunits adopt a filamentous arrangement that may resemble the biologically relevant assembly. Supporting experiments confirm that one pilin subunit is connected to the next by an isopeptide bond between a lysine side chain and the Leu-Pro-X-Thr-Gly motif.

In addition to illuminating the isopeptide bond that links the subunits, the structure reveals important unanticipated features. It shows that the major pilin subunit consists of two similar β-sheet domains, each of which is stabilized by its own internal isopeptide bond (see the figure, middle). Kang *et al.* identify likely mechanisms that allow these bonds to be formed autocatalytically, drawing parallels to the isopeptide bonds that stabilize the HK97 viral capsid (*9*).

The importance of the intrasubunit bonds is evident in hindsight. The pili of Gram-negative bacteria have two structural advantages relative to those of Gram-positive bacteria. In Gram-negative bacteria, the pili are relatively thick, being either tubular or rod-shaped, with cross sections of three to six subunits (*10*). Also, typical Gram-negative organisms can stabilize individual subunits with disulfide bonds. In contrast, current data suggest that typical Gram-positive pili are only one subunit thick. To stabilize such a thin filament, the isopeptide bonds between subunits may need to be supplemented by further covalent bonds within the subunits.

If the purpose of these bonds were simply to stabilize the individually folded protein domains, this might have been accomplished by bonds at various positions. Instead, the bonds occur at positions that lead to a nearly linear chain of covalent connectivity along the entire pilus (see the figure, middle).

In view of its function, this design is ingenious. The role of the pilus in maintaining host cell attachment implies that it must endure considerable tensile forces. Given the placement of the isopeptide bonds—both between

The authors are in the Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA. E-mail: yeates@mbi.ucla.edu

and within subunits—tensile forces along the pilus should exert little if any unfolding forces on the separate pilin domains. Kang *et al.* show that several other cell surface adhesion proteins have similar isopeptide bonds (see the figure, right) that were overlooked in previous structural studies.

The structural framework for Gram-positive pili elucidated by Kang *et al.* opens up new lines of investigation into proteins important in bacterial pathogenesis. It also advances our general understanding of protein assemblies. Filamentous protein assemblies occur widely in nature, as well as in human pathologies, but our understanding of them is still seriously incomplete. This new research provides a valuable addition to the short list of filamentous assemblies that have been characterized in atomic detail.

References

1. J. R. Scott, D. Zahner, *Mol. Microbiol.* **62**, 320 (2006).
2. J. L. Telford *et al.*, *Nat. Rev. Microbiol.* **4**, 509 (2006).
3. H. Ton-That, O. Schneewind, *Trends Microbiol.* **12**, 228 (2004).
4. H. J. Kang, F. Coulibaly, F. Clow, T. Proft, E. N. Baker, *Science* **318**, 1625 (2007).
5. H. Ton-That, O. Schneewind, *Mol. Microbiol.* **50**, 1429 (2003).
6. A. Swaminathan *et al.*, *Mol. Microbiol.* **66**, 961 (2007).
7. A. G. Manetti *et al.*, *Mol. Microbiol.* **64**, 968 (2007).
8. M. Mora *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15641 (2005).
9. W. R. Wikoff *et al.*, *Science* **289**, 2129 (2000).
10. L. Craig, M. E. Pique, J. A. Tainer, *Nat. Rev. Microbiol.* **2**, 363 (2004).

10.1126/science.1151398

PHYSICS

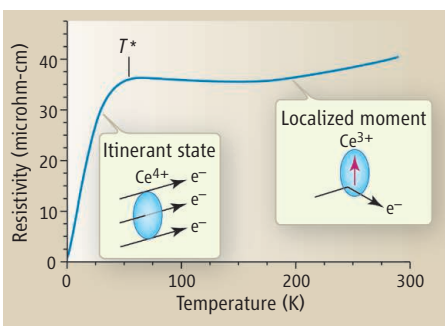
A Whiff of Chemistry in Heavy Electron Physics

Calculations show how electrons in a rare earth compound develop their unusual low-temperature properties.

Zachary Fisk

Metallic compounds of the rare earth elements are a wonderful arena for studies of magnetism. The electrons responsible for the magnetic properties of these elements (the so-called 4f electrons) are localized within the core electron states and carry magnetic moments that can be simply derived from rules introduced in a first-year chemistry course. However, cerium, at the start of the rare earth series, is an exception because its 4f level is less well localized in the atomic core. In this case, magnetic measurements often reveal a distinctive behavior in which the localized magnetic moments evolve into itinerant behavior similar to that of metallic electrons (see the figure). This evolution from local to itinerant behavior can have unusual consequences, and on page 1615 of this issue, Shim *et al.* (*1*) report their state-of-the-art computational studies of how electrons behave in such materials. Their calculations can follow the evolution of a disordered high-temperature electronic state into a coherent low-temperature electronic state with strongly altered properties.

As the magnetic moments change from local to itinerant behavior, some of their properties are transferred to the conduction electrons. This is reflected in the conduction electrons moving much more slowly in the lattice, effectively possessing an enhanced mass one to three orders of magnitude larger than that of free electrons, and phase transitions of this heavy-mass sea of electrons are possible at low temperature. It was an astonishing discovery in 1979 by Steglich *et al.* (*2*) that superconductivity could occur in such a system. Equally surprising has been the slowly developing conviction, as more examples have been discovered, that this type of superconductivity is always associated with a magnetically ordered state that can be induced by a small variation of some parameter of the system, be it chemical, external magnetic field, or pressure.



Developing coherence. Graph of electrical resistivity of $CeIrIn_5$ as a function of temperature. At high temperatures, the electrons are scattered from magnetic Ce^{3+} moments in the material, causing high resistivity. Below T^*, the moments "dissolve" by giving up an electron into the sea of electrons, which gain coherence and form a heavy electron fluid with reduced scattering.

$CeIrIn_5$ is such a superconductor, belonging to a series that includes superconducting cobalt and antiferromagnetic rhodium homologs (*3*–*5*). Researchers are especially interested in this set of compounds because of the close proximity of magnetic and superconducting ground states coupled with the relative ease with which single crystals of high quality can be grown. Calculations of electronic structure for these materials do not have the resolution to suggest why two are superconductors and one is an antiferromagnet, yet this lies at the heart of what experimenters in the study of such materials would like to understand.

Shim *et al.* bring the latest computational techniques to bear on this problem in their study of the iridium homolog. These techniques have been developed in the past decade to address strong electron-electron interactions, which are particularly difficult to handle in conventional band calculations and which dominate the low-temperature physics here. Remarkably, Shim *et al.* are able to track the local-itinerant evolution with temperature and match their calculations with temperature-dependent optical spectroscopic data.

The local-itinerant transition corresponds to the development of coherence: In a chemically ordered lattice, one expects phase-coherent Bloch electronic states (i.e., the wave functions of an electron in a periodic potential) having the symmetry of the crystal lattice. These are the states computed in typical band-structure calculations. The calculations of Shim *et al.* follow how these Bloch states actually develop with decreasing temperature, supporting a two-fluid picture (*6*) that has been used to model the evolving low-temperature itinerant state, similar in ways to a gas/liquid transition. As a result, a direct comparison can now be made between the computed heavy-mass evolution with temperature and experimental data (*7*), which is an impressive achievement. A further interesting result comes from the ability of Shim *et al.* to isolate specific atom-atom interactions. They find the strongest cerium-indium interaction with the indium near neighbors of iridium, giving a hint of why iridium may be different from

The author is in the Department of Physics and Astronomy, University of California at Irvine, Irvine, CA 92697–4575, USA. E-mail: zfisk@uci.edu

CREDIT: ADAPTED FROM C. PETROVIC *ET AL.*

the rhodium and different from the cobalt homolog (which has a five times higher superconducting transition temperature T_c).

Why is this finding important? The class of heavy-mass materials based on cerium and certain other rare earth and actinide elements are the simplest examples of so-called highly correlated electron materials—simplest because chemical bonding effects associated with the local-itinerant evolution are minimal. The interesting physics of these materials is controlled by the temperature scale T^*, the so-called coherence scale below which the phase-coherent Bloch electron states develop. T^* can be simply determined from specific-heat measurements. This same temperature scale falls out naturally in the two-fluid empirical description of heavy-mass systems and is also clearly evident in nuclear magnetic resonance and magnetic susceptibility measurements (*8*). It is an interesting possibility that T^* may set the scale for the superconducting T_c of the heavy-mass superconductors: T_c is ~1/20 of T^*, similar to the relation of T_c and the lattice vibration temperature scale relevant to conventional superconductors. This and other similarities suggest some physics common to both heavy-mass superconductors and high-T_c cuprates.

We know essentially nothing about what determines T^*, yet this is the parameter of central importance to the physics of these materials. A hint of an effective route toward understanding what determines T^* and how to reach higher T^* comes from the work of Shim *et al*. The longer view is that what we learn from the heavy-mass materials may reach into certain transition metal materials. Materials research in cerium and related compounds is still in the hunter-gatherer stage. With the help of these powerful new computational methods may come advancement to effective materials husbandry and an understanding of the chemistry that results in these physical effects.

References

1. J. H. Shim *et al*., *Science* **318**, 1615 (2007); published online 1 November 2007 (10.1126/science.1149064).
2. F. Steglich *et al*., *Phys. Rev. Lett*. **43**, 1892 (1979).
3. H. Hegger *et al*., *Phys. Rev. Lett*. **84**, 4986 (2000).
4. C. Petrovic *et al*., *Europhys. Lett*. **53,** 354 (2001).
5. C. Petrovic *et al*., *J. Phys. Condens. Matt*. **13**, L337 (2001).
6. S. Nakatsuji *et al*., *Phys. Rev. Lett*. **92**, 016401 (2004).
7. Y. Yang, D. Pines, http://arxiv.org/abs/0711.0789.
8. N. J. Curro *et al*., *Phys. Rev*. **B 64**, 180514 (2001).

10.1126/science.1151945

ASTRONOMY

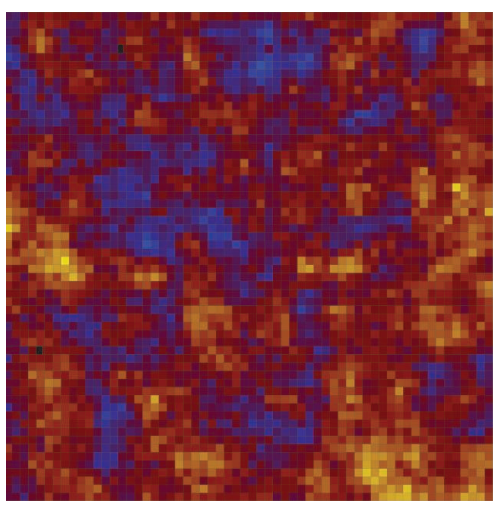
A Texture in the Sky?

The cosmic microwave background may contain the echoes of a defect in spacetime that occurred just after the Big Bang.

Robert Brandenberger

Since the 1960s, we have known that the sky is filled with the afterglow of the Big Bang. Imprinted on this cosmic microwave background (CMB) may be clues to the earliest moments of the universe. On page 1612 of this issue, Cruz *et al.* (*1*) argue that a cold spot observed in recent observations of the microwave sky could be caused by a kind of ancient cosmic ripple called a texture. If confirmed, this result would provide a direct link between observations and physics at energies close to the Planck scale, where quantum effects are believed to become important in gravity. This energy scale far exceeds present and future terrestrial experiments, but understanding physical processes at this scale may help us combine two phenomena—quantum mechanics and general relativity—that so far have resisted reconciliation.

According to the current standard paradigm of early universe cosmology, known as the "inflationary universe scenario" (*2*, *3*), the inhomogeneities that we observe today in the distribution of galaxies on large scales and the anisotropies in the CMB are the result of quantum fluctuations at very early times, when the universe was dominated by an unknown field that caused an accelerated expansion of space. The theory that describes how these fluctuations evolved from the earliest moments of the universe to the present time was worked out in the early 1980s (*4*) [see, e.g., (*5*) for a comprehensive review] and predicted the detailed nature of the angular distribution of the CMB, a prediction that was spectacularly confirmed in recent CMB experiments (*6*, *7*). Despite this phenomenological success, a number of cosmologists have been puzzling over some important conceptual problems in our current models of inflationary cosmology [see, e.g., (*8*) for a discussion of some of these issues] and wondering about alternatives to inflation.



Defects in the universe. Map of the CMB temperature in a 15° by 15° portion of the sky in a model with both Gaussian noise and cosmic strings (10 arc-minute resolution). The string tension is set by a scale comparable to the energy scale of the texture of Cruz *et al.* (*1*). Colors indicate temperatures of the CMB (blue for colder, red and yellow for hotter). The relative temperature difference between hot and cold spots are on the order of 10^{-5}. Note the edges across which the temperature jumps.

CREDIT: ADAPTED FROM AMSEL *ET AL.* (*18*)

Whereas the physics of inflation is rather mysterious, physicists know for sure that matter undergoes phase transitions as it is cooled. For example, as water cools it turns to ice. Depending on the properties of the matter, topological defects (such as defects in an ice crystal) may form. The key point for applications of defects to cosmology is that, if the microphysics (that is, the underlying laws that govern particles and forces) allows for defects, then such defects in the alignment of matter inevitably formed during a phase transition in the early universe (*9*).

Topological defects come in various types: pointlike defects (monopoles), linear defects (cosmic strings), planar defects (domain walls), and spherical collapsing defects, which are called "textures" in the cosmology literature. Of these defects, only strings and textures are viable (models that predict monopoles or domain walls are ruled out by observations). In the mid- to late 1980s, models in which the energy associated with topological defects were responsible for the generation of structure in the universe were popular [see, e.g., (*10*) for a comprehensive review]. However, the detailed measurements of the angular distribution of CMB anisotropies did

The author is in the Physics Department, McGill University, 3600 University Street, Montreal, Canada, H3A 2T8. E-mail: rhb@physics.mcgill.ca

not confirm the predictions of topological defect models (*11*–*13*), and interest in these models decreased dramatically.

These measurements did not rule out topological defects, however. Rather, they only implied that defects could not be the dominant source of CMB anisotropies. There has, in fact, been a lot of recent interest in the possibility that cosmic strings produced by superstring theory might be seen in the sky [see, e.g., (*14*) for a review] and might contribute a fraction less than 10% to the overall anisotropies in the CMB sky. The main signature of cosmic strings in the sky would be lines across which the temperature jumps by an amount proportional to the tension of the string (see the figure) (*15*). The imprint of cosmic textures would be a distribution of hot and cold spots in the sky whose angular diameter depends on the time when the textures collapse (spots of between 1° and 10° are expected, with the smaller spots being more numerous) (*16*, *17*). In the case of textures, the temperature deviation of the hot or cold spots from the average temperature depends on the energy scale of the texture, which in turn tells us the temperature at which the phase transition that generated the textures occurred. An important point is that these features of topological defects in the CMB maps, which can easily be identified when analyzing the actual maps, get washed out when analyzing the data in the usual way, namely decomposing the maps into angular components and plotting the amplitude of each angular component (or, in more technical terms, when calculating the angular power spectrum).

Since the release of the CMB data from the Wilkinson Microwave Anisotropy Probe (WMAP) (*7*), some apparent anomalies in the temperature maps have been pointed out. Cruz *et al.* focus on a cold spot of angular size roughly 5° that had previously been identified. They performed two sets of numerical simulations of sky maps, one based on fluctuations from inflation only ("Gaussian maps"), the other assuming Gaussian maps (with reduced amplitude) plus a temperature template produced by a cosmic texture. Based on a Bayesian analysis, the authors find a probability ratio of 2.5:1 favoring the texture plus Gaussian model over the Gaussian model. The authors discuss follow-up tests with which to further test the hypothesis that the observed cold spot is due to a texture.

If confirmed, the identification of a cosmic texture in the sky will have provided us with good evidence that a phase transition in matter occurred at an energy of roughly 10^{16} GeV, many orders of magnitude higher than energy scales that can be reached in terrestrial experiments. The energy scale involved in the candidate texture is close to the elusive Planck scale, an energy where quantum gravity becomes important. A lesson to be learned from this work is that Planck-scale physics may well be testable in the very near future in cosmological observations.

References

1. M. Cruz *et al.*, *Science* **318**, 1612 (2007); published online 25 October 2007 (10.1126/science.1148694).
2. A. H. Guth, in *Measuring and Modeling the Universe*, W. L. Freedman, Ed. (Carnegie Observatories, Pasadena, CA, 2004), pp. 31–52.
3. A. Linde, *Particle Physics and Inflationary Cosmology* (Harwood, Chur, Switzerland, 1990).
4. V. F. Mukhanov, G. V. Chibisov, *JETP Lett.* **33**, 532 (1981).
5. V. F. Mukhanov, *et al.*, *Phys. Rept.* **215**, 203 (1992).
6. C. B. Netterfield *et al.* (Boomerang Collaboration), *Astrophys. J.* **571**, 604 (2002), www.arxiv.org/abs/astro-ph/0104460.
7. C. L. Bennett *et al.* (WMAP Collaboration), *Astrophys. J. Suppl.* **148**, 1 (2003).
8. R. H. Brandenberger, www.arxiv.org/abs/hep-ph/9910410.
9. T. W. B. Kibble, *J. Phys. A* **9**, 1387 (1976).
10. A. Vilenkin, E. P. S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge Univ. Press, Cambridge, 1994).
11. U. L. Pen *et al.*, *Phys. Rev. Lett.* **79**, 1611 (1997).
12. R. Durrer *et al.*, *Phys. Rev. D* **59**, 123005 (1999).
13. A. Albrech *et al.*, *Phys. Rev. Lett.* **79**, 4736 (1997).
14. J. Polchinski, www.arxiv.org/abs/hep-th/0412244.
15. N. Kaiser, A. Stebbins, *Nature* **310**, 391 (1984).
16. N. Turok, *Phys. Rev. Lett.* **63**, 2625 (1989).
17. N. Turok, D. Spergel, *Phys. Rev. Lett.* **64**, 2736 (1990).
18. S. Amsel *et al.*, www.arxiv.org/abs/0709.0982.

10.1126/science.1151994

PLANT SCIENCE

The Power of the Pyramid

Insects may be less likely to develop resistance when crops express proteins that have been modified to overcome resistance to other toxins.

William J. Moar and Konasale J. Anilkumar

The use of crops that are genetically engineered to produce *Bacillus thuringiensis* (*Bt*) toxins has risen rapidly to more than 32 million hectares in 2006, resulting in substantially reduced use of insecticides and increased grower profit (*1*). However, with the increased use of *Bt* crops, such as corn and cotton (see the figure), comes the threat that target pests may develop resistance to these toxins. To date, there have been no reports of *Bt* resistance occurring in field populations of insects during the 11 years that *Bt* crops have been commercialized (*2*). Have we just been lucky, or have there been safeguards built in to the plants, and their use, to delay resistance development? To help answer this question, we need to know more about how *Bt* toxins kill insects. On page 1640 in this issue, Soberón *et al*. (*3*) provide a breakthrough in understanding the mode of action of *Bt* toxins. The molecular details may be relevant for producing modified *Bt* proteins that are still toxic to target insects, yet do not drastically affect the stability or host range of *Bt* proteins now in use, thus hopefully avoiding regulatory hurdles for registering novel insecticidal compounds.

Pink bollworm (*Pectinophora gossypiella*) European corn borer (*Ostrinia nubilalis*)

Crop pests. Insects may develop resistance to plants that are genetically engineered to express *Bt* toxins or to insecticides containing *Bt* toxins. Pink bollworm (*Pectinophora gossypiella*) attack cotton (**left**) and the European corn borer (*Ostrinia nubilalis*) attack corn (**right**).

PHOTO CREDITS: (PINK BOLLWORM) P. GREB/USDA AGRICULTURAL RESEARCH SERVICE/BUGWOOD.ORG; (EUROPEAN CORN BORER) K. WELLER/USDA AGRICULTURAL RESEARCH SERVICE/BUGWOOD.ORG

In the United States, resistance-management strategies are mandated before *Bt* crops can be grown (*4*). First-generation *Bt* crops primarily use the "high dose plus refugia" option, whereby plants are engineered to express enough *Bt* protein to kill all target insects except for the very rare resistant individuals. These rare insects presumably have two copies (alleles) of the gene(s) that confer resistance. As for insects with recessive alleles for such genes, they are thought to be "diluted out" by susceptible insects from the refugia. However, even this strategy is not

The authors are in the Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, USA. E-mail: moarwil@auburn.edu

expected to last forever, although the 11 years of utility (and still counting) for first-generation *Bt* crops already is remarkable (*2*).

One solution has been to engineer crops that express at least two toxic compounds that act independently, so that resistance to one does not confer resistance to the other. This approach, called gene pyramiding, became a commercial reality in 2003 with the introduction of Bollgard II, a transgenic cotton plant that expresses the original *Bt* protein, Cry1Ac, and a second *Bt* protein, Cry2Ab. The two proteins act independently in that they bind to different receptors in the insect's midgut. Of course, additional compounds for pyramiding are needed, but finding them is difficult. Each candidate must be encoded by a single gene (for transgenic plant development), must be toxic to the target pest, and must demonstrate a different mechanism of action from *Bt* toxin(s) already in the plant. Beyond those criteria, if the compound is novel, it must go through extensive regulatory testing.

So far, there are relatively few candidates for gene pyramiding in *Bt* cotton and corn for controlling target lepidopteran pests (*5*). One promising way to improve this is to determine how target pests develop resistance to specific toxins, and then modify these toxins so that resistance must occur in another manner. This would plausibly increase the time for resistance to develop and increase the life expectancy of insect-resistant *Bt* crops.

Soberón *et al*. confirm that active *Bt* toxins require additional enzyme cleavage before toxicity can occur (*6*, *7*). Moreover, in the absence of a functional toxin receptor (cadherin) to properly bind the active forms of *Bt* toxin, cleavage does not occur. More importantly, Sóberon *et al*. constructed modified *Bt* proteins (Cry1AbMod and Cry1AcMod) that were artificially "cleaved." These modified proteins were still toxic to insects that no longer expressed functional cadherin proteins, as well as to insects that were already resistant to native forms of the toxins (because they expressed mutated cadherin proteins that do not bind toxins). Thus, these modified proteins could potentially bolster a gene pyramiding scheme for delaying the development of insect resistance in crops.

One primary question arises: Can Cry1 AbMod and Cry1AcMod be expressed at high levels in crops, and control target pests that have become resistant (due to mutations in cadherin) to the original *Bt* proteins they were derived from? Soberón *et al*. suggest that one possible reason Cry1AbMod and Cry1AcMod were slightly less toxic than native *Bt* toxins is that they could be less stable in the insect's midgut. If so, there might be concerns that they won't be stable in plant cells and tissues, already one constraint in maintaining high *Bt* protein expression throughout the growing season (*8*). The knowledge that *Bt* toxins require additional enzymatic processing after binding to cadherin could hopefully lead to the design of a *Bt* protein that is specifically prone to enzymes in the midgut (without requiring cadherin binding), but not more susceptible to host plant enzymes.

Although *Bt* resistance has been the primary environmental concern with *Bt* crops in the United States, there are few laboratory-generated *Bt*-resistant insect model systems, and none have evolved from *Bt* crops. There are even fewer *Bt*-resistant model insects that have been selected for resistance to one particular *Bt* protein. For Cry1Ac, there are two such *Bt*-resistant insect models available in the United States—*Heliothis virescens* and *Pectinophora gossypiella*, both pests of cotton (*9*–*11*). Interestingly, and in support of the concept proposed by Soberón *et al*., both *Bt*-resistant insects (and a third in China) (*12*) express altered cadherin. As more *Bt*-resistant insect model systems become available, it will be necessary to determine how universal mutations in cadherin are, and whether alterations in non-cadherin binding regions of *Bt* proteins could be made to delay other potential mechanisms of resistance.

As we continue to alter *Bt* proteins from their natural structure and composition, the question arises as to whether the selectivity or host range of these modified proteins will be altered as well. Clearly, this will need to be addressed, but the concept of designing *Bt* proteins to pyramid with other compounds to delay *Bt* resistance warrants further investigation.

References

1. C. James, "Brief 35: Global Status of Commercialized Biotech/GM Crops: 2006" (International Service for the Acquisition of Agri-Biotech Applications, Ithaca, NY, 2006).
2. M. I. Ali, R. G. Luttrell, *J. Econ. Entomol.* **100**, 921 (2007).
3. M. Soberón *et al*., *Science* **318**, 1640 (2007); published online 1 November 2007 (10.1126/science.1146453).
4. U.S. Environmental Protection Agency, www.epa.gov/pesticides/biopesticides/pips/bt_brad.htm (2001).
5. W. J. Moar, *Nat. Biotechnol.* **21**, 1152 (2003).
6. I. Gómez, J. Sánchez, R. Miranda, A. Bravo, M. Soberón, *FEBS Lett.* **513**, 242 (2002).
7. N. Jiménez-Juárez *et al.*, *J. Biol. Chem.* **282**, 21222 (2007).
8. Y. Gao *et al*., *J. Agric. Food Chem*. **54**, 829 (2006).
9. L. J. Gahan, F. Gould, D. G. Heckel, *Science* **293**, 857 (2001).
10. R. Y. Xie *et al*., *J. Biol. Chem.* **280**, 8416 (2005).
11. S. Morin *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5004 (2003).
12. X. Xu, L. Yu, Y. Wu, *Appl. Environ. Microbiol.* **71**, 948 (2005).

Published online 1 November 2007;
10.1126/science.1151313
Include this information when citing this paper.

ECOLOGY

How Do Roots Interact?

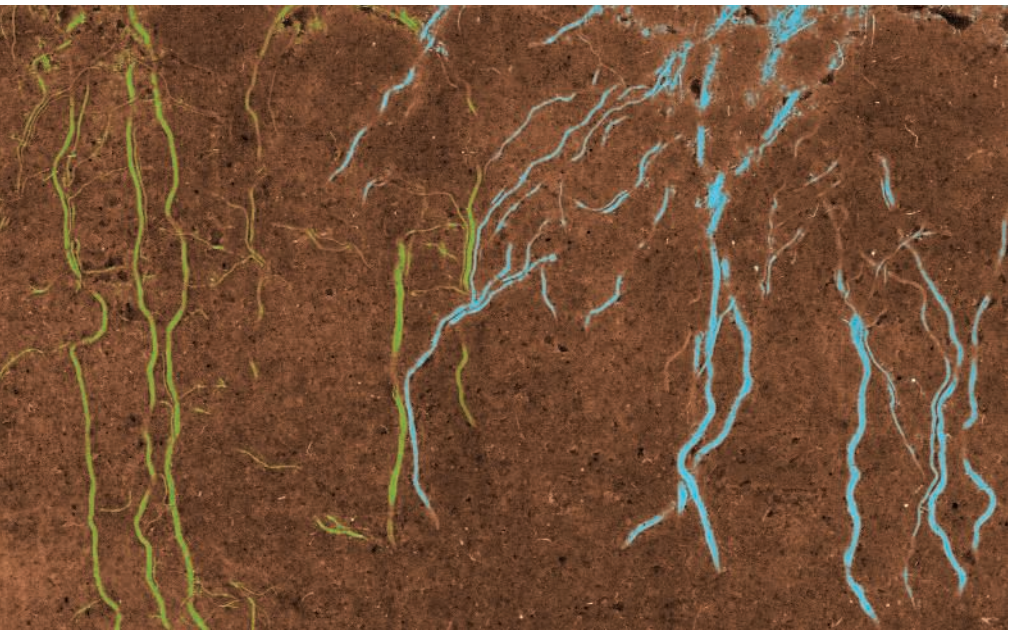
Hans de Kroon

Plant roots recognize and respond to the identities of their neighbors.

The main function of plant roots is the acquisition of mineral nutrients and water from the soil. Roots do not encounter these belowground resources passively, but actively forage for nutrient hot spots (*1*) and avoid patches where root densities of competing neighbors are high (*2*). These responses can be driven by local nutrient concentrations in the soil (*3*). However, it is becoming increasingly clear that the world underground is even more complex and that elaborate root interaction mechanisms are at work.

Several studies have shown that roots respond to neighboring roots in a very specific manner that depends on the identity of the neighbor (*4*–*6*). Root extension tends to be greater when roots grow into substrate containing "nonself" roots of a genetically different individual or a detached plant with the same genotype than when "self" roots of the same (physiological and genetic) individual are encountered. Dudley and File have recently shown that plants of the Great Lakes Sea Rocket (*Cakile edentula*) invested more biomass in fine roots when they competed with unrelated individuals than when they competed with siblings (*7*). This is one of the few cases (*6*, *8*) in which root behavior has been shown to depend solely on the genetic identity of competing roots. Depending on the species, genotypic or physiological recognition processes appear to be involved in these root interactions.

The author is in the Department of Experimental Plant Ecology, Institute for Water and Wetland Research, Radboud University, 6525 ED Nijmegen, the Netherlands. E-mail: H.deKroon@science.ru.nl



CREDIT: M. SEMCHENKO/UNIVERSITY OF SUSSEX, UK

Neighbor contest underground. In this image from a root observation chamber, roots from wild strawberry (colored blue) approach ground ivy roots (colored green). Strawberry root growth is stimulated by ground ivy roots, whereas ground ivy root growth is inhibited by strawberry roots (*9*). Such species-specific root recognition mechanisms may affect community dynamics.

To date, root recognition studies have focused almost exclusively on competition between individuals of the same species. In another recent study, Semchenko *et al.* (*9*) add a new dimension by examining the interactions between two different species from the same plant community: ground ivy (*Glechoma hederacea*) and wild strawberry (*Fragaria vesca*). Contrary to expectation (*2*), when competing in trays, wild strawberry plants produced as much root mass into the ground ivy neighborhood as in a comparable soil volume in the opposite direction away from the competitor (see the figure). In contrast, ground ivy roots avoided the wild strawberry neighborhood. For both species, root development was similar when a plant was confronted with competition from the roots of the same plant, a detached plant of the same genotype, or a different genotype of the same species (*9*), suggesting that physiological (*5*) or genetic (*7*) recognition is not necessarily a general phenomenon.

What is the ecological and evolutionary importance of these idiosyncratic root interactions? One persistent hypothesis is that restraining root development in a "self" neighborhood saves resources and has evolved because genotypes can invest the saved resources in enhanced reproduction (*4*, *5*, *7*). However, it has proved very difficult to confirm this hypothesis experimentally. A number of studies did show lower plant reproduction associated with nonself root growth stimulation, but these results have been criticized due to pot size artifacts in the experimental design (*10*). Valid tests of this hypothesis are still needed, but alternative explanations should also be considered.

One straightforward alternative hypothesis would be that greater root growth in a nonself neighborhood enhances fitness and is selected for. The costs of making more roots may be offset by the benefits of elevated resource acquisition and plant growth. Indeed, especially in the initial growth phase, plants with elevated root extension in a nonself neighborhood tend to be larger than plants with restricted root growth in a self neighborhood (*4*, *7*, *10*). If not constrained to a small pot volume relative to mature plant size, the larger root mass will eventually exploit a larger part of the contested soil resources, resulting in enhanced competitive ability (*6*, *11*). Larger plants also have higher survival and fecundity (that is, higher Darwinian fitness) (*11*).

If these sophisticated root interactions operate in agro-ecosystems, one would predict that root production would be enhanced in genetically diverse crops or in intercropping where different species interact. Indeed, crops in mixtures have been shown to produce more roots and to explore a larger soil volume than in monoculture (*12*). Consistent with the hypothesis that greater root growth is beneficial in these interactions, these intercropping systems also had a higher yield. In a recent paper on maize–faba bean intercropping, Li *et al.* (*13*) explain this overyielding by the release of organic acids by the faba bean roots. The resulting acidification of the soil enhances the mobilization of phosphorus, which benefits faba bean plants and other plant species whose roots intermingle.

These results indicate that facilitation among species underlies root growth stimulation, although specific recognition could still play a role. It seems unlikely that this particular facilitative mechanism can explain all the root responses seen in other systems that are less phosphorus-limited. Future studies must unravel the role of such facilitative effects in relation to other root interaction mechanisms, including nutrient foraging responses, recognition mechanisms, and the many other positive and negative plant-plant interactions mediated by root exudates (*14*).

Natural communities have a more complicated structure, with plant species distributed in more or less aggregated patterns that may depend on clonal growth or limited seed dispersal. Differences in root behavior may be associated with differences in aggregation. Semchenko *et al*. (*9*) suggest that the more clumped *Glechoma* creates its own territory (*15*), where it pays off to avoid competition with neighboring species, whereas the more spread-out *Fragaria* has more interspecific contacts and challenges the contest. Differential root interactions may thus consolidate spatial patterning in communities, which in turn profoundly affects biodiversity and community dynamics (*16*).

Thus, a range of root responses may influence plant performance in natural and agricultural ecosystems and may affect the interactions and distributions of populations and species. Many details of these processes and their effects remain unknown and merit full investigation.

References and Notes

1. A. Hodge, *New Phytologist* **162**, 9 (2004).
2. M. Gersani *et al*., *Evol. Ecol.* **12**, 223 (1998).
3. H. M. Zhang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6529 (1999).
4. G. G. Maina *et al*., *Plant Ecol.* **160**, 235 (2002).
5. O. Falik *et al*., *J. Ecol.* **91**, 525 (2003).
6. H. de Kroon, L. Mommer, A. Nishiwaki, in *Root Ecology*, H. de Kroon, E. J. W. Visser, Eds. (Springer, Berlin, 2003), pp. 215–234.
7. S. A. Dudley, A. L. File, *Biol. Lett.* **3**, 435 (2007).
8. B. E. Mahall, R. M. Callaway, *Am. J. Botany* **83**, 93 (1996).
9. M. Semchenko *et al*., *New Phytologist* **176**, 644 (2007).
10. L. Hess, H. de Kroon, *J. Ecol.* **95** 241 (2007).
11. J. Weiner, *Trends Ecol. Evol.* **5**, 360 (1990).
12. L. Li *et al.*, *Oecologia* **147**, 280 (2006).
13. L. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11192 (2007).
14. H. P. Bais *et al*., *Ann. Rev. Plant Biol.* **57**, 233 (2006).
15. H. J. Schenk *et al*., *Adv. Ecol. Res.* **28**, 145 (1999).
16. L. A. Turnbull *et al*., *J. Ecol.* **95**, 79 (2007).

17. I am grateful to H. During, L. Hess, M. Hutchings, S. Kembel, M. Semchenko, and E. Visser for comments and discussion.

10.1126/science.1150726

RETROSPECTIVE

Arthur Kornberg (1918–2007)

One of the greatest biochemists of the 20th century and a "lover of enzymes" is remembered.

Paul Berg and I. R. Lehman

Arthur Kornberg, who had a life-long love affair with enzymes, died on 26 October surrounded by his family and mourned by his extended family of students and colleagues. It is not surprising that only 2 weeks before, he had been actively summarizing decades of work on polyphosphate for a review article.

In his autobiography, *For the Love of Enzymes: The Odyssey of a Biochemist*, Arthur described his entry into science and evolution from clinician to nutritionist to biochemist. He was born in Brooklyn, New York, on 3 March 1918, the son of parents who had emigrated from Eastern Europe and ran a small hardware store. He majored in chemistry and biology at City College of New York. After receiving an M.D. from the University of Rochester, he entered the U.S. Public Health Service and served briefly as a ship's doctor during World War II.

But he was enormously influenced by the great biochemists of the 1930s and 1940s—F. G. Hopkins, Otto Warburg, and Otto Meyerhoff—and then by Carl and Gerti Cori and Severo Ochoa, in whose laboratories he worked. His career in medicine changed to research in biochemistry. By exploiting the power of enzyme purification to reconstitute biochemical pathways—what he called "the hammer of enzyme purification"—Arthur undertook the formidable problem of synthesizing DNA. While at Washington University from 1953 to 1959, he discovered the first DNA polymerase and established DNA synthesis as a template-driven process, for which he shared the 1959 Nobel Prize in Physiology or Medicine.

Together with talented students and postdoctoral fellows, Arthur accomplished what some described as the creation of life in a test tube—recreating a bacteriophage chromosome with purified enzymes, substrates, and cofactors. Unfazed by that success, his laboratory reconstituted the complex process of bacterial chromosome replication in vitro. That monumental achievement influenced a generation of biochemists to undertake problems seemingly as intractable as gene expression, signal transduction, and intracellular protein transport. The ability to clone, amplify, and sequence genes, and the "biological revolution" that followed, were possible largely because of the enzymes that emerged from Arthur's pioneering work.

In the early 1990s, Arthur turned from DNA replication to the study of polyphosphate (polyP), a molecule that had intrigued him since the 1950s when he and his first wife Sylvy isolated polyphosphate kinase (PPK), which synthesizes polyP. His studies of polyP and PPK, which, as he put it, "disinterred a molecular fossil," led to the discovery of polyP's role in bacterial growth and survival, quorum sensing, biofilm formation, and virulence. He was convinced that future work would reveal the clinical importance of polyP in microbial infections.

Beyond his scientific achievements, Arthur's considerable expository gifts and the ability to project his ideas are exemplified by his superb 1980 textbook, *DNA Replication*, which educated a generation of molecular biologists. Fred Sanger conceived the "dideoxy" DNA sequencing method while reading the chapter on DNA polymerase I. In *The Golden Helix: Inside Biotech Ventures*, Arthur drew on his experience as a founder of the DNAX Research Institute of Molecular and Cellular Biology to provide a unique perspective on biotechnology. His last book, *Germ Stories*, a collection of poems for children, reveals the wonders and hazards of the microbial world.

Arthur's contributions to science did not go unrecognized. In addition to the Nobel Prize, he received the National Medal of Science, the Cosmos Club Award, and the Gairdner Foundation Award, among others. He served as president of the American Society of Biological Chemists, was elected to membership in the U.S. National Academy of Sciences, the American Academy of Arts and Sciences, and the American Philosophical Society, and was a Foreign Member of the British Royal Society. He was also awarded honorary doctorates from 12 universities.

Arthur revealed his gift as a leader by organizing the Enzyme and Metabolism Section of the National Institute of Arthritis and Metabolic Diseases. He assembled an outstanding Department of Microbiology at the Washington University School of Medicine in St. Louis, Missouri, as well as the Department of Biochemistry at Stanford. We accompanied him to Stanford along with Melvin Cohn, David Hogness, Dale Kaiser, and Robert Baldwin. Five of the six faculty members who accompanied him from St. Louis in 1959 have remained at Stanford, a tribute to Arthur's leadership.

In an unusual and much admired arrangement initiated by Arthur at Washington University and maintained at Stanford, the department's graduate students and postdoctoral fellows were provided available space in common laboratories. This encouraged research groups to interact and share reagents and methods, practices that greatly facilitated development of recombinant DNA technology at Stanford.

Both of us knew Arthur for more than 50 years, from the time we joined his laboratory at Washington University as postdoctoral fellows. But our relationships with him went beyond that of student and mentor. We were embraced as members of his family and shared many special occasions and achievements that they celebrated. Arthur's style of doing science, his passion for experimentation rather than theory, and excitement about discovery inspired us. We remember the late-night calls inquiring how our experiments had fared. He was a serious and superb teacher and a generous and compassionate leader. The success of the faculties he assembled attests to his gift of forsaking the limelight and encouraging his colleagues to flourish on their own. Above all, Arthur was devoted to his students and colleagues and fiercely loyal to his family and friends. Perhaps Arthur's greatest legacy, and certainly the one of which he was most proud, was his extraordinary family of three sons and eight grandchildren. We will miss him greatly.

The authors are at Stanford University, Stanford, CA 94305, USA. E-mail: pberg@cmgm.stanford.edu

10.1126/science.1152989

CORRECTED 21 DECEMBER 2007; SEE LAST PAGE

GE PRIZE-WINNING ESSAY

Why Old World Monkeys Are Resistant to HIV-1

Premature disassembly of the HIV-1 capsid is caused by the rhesus monkey protein, TRIM5α, and a single amino acid change in human TRIM5α confers similar anti–HIV-1 activity.

Matthew Stremlau

GE Healthcare and *Science* are pleased to present the prize-winning essay by Matt Stremlau, a regional winner for North America who is the Grand Prize winner of the GE & *Science* Prize for Young Life Scientists.

GE & Science Prize for Young Life Scientists

Humans have been exposed to retroviruses for millions of years. Indeed, a significant portion of our genome consists of endogenous retroviruses—reminders of our vulnerability to past infections. The HIV/AIDS epidemic, which began nearly a century ago when simian immunodeficiency virus (SIV) passed from chimpanzees into a human host, is the latest episode in the long-standing coevolutionary struggle between retroviruses and their hosts.

Human immunodeficiency virus type 1 (HIV-1) causes AIDS in humans, and to a lesser extent, in chimpanzees (*1*, *2*). However, not long after the discovery of HIV-1, scientists realized that certain primate species were resistant to HIV-1 infection. In particular, monkeys from Africa and Asia, referred to as Old World monkeys, could not be infected with HIV-1 and did not develop AIDS (*3*). This discovery brought both excitement and frustration. The block to HIV-1 replication in Old World monkey cells hindered efforts to develop an animal model for testing drugs and vaccines. On the other hand, Old World monkeys had evolved for millions of years in Africa—the epicenter of the current HIV-1 epidemic. Perhaps exposure to past HIV-1–like epidemics led to the emergence of an antiviral defense that protected them against HIV-1.

Determining the cause of HIV-1 resistance in Old World monkey cells stymied HIV researchers for nearly two decades. An early view was that the block resulted from expression of an incompatible receptor on the surface of Old World monkey cells. However, identification of the HIV-1 co-receptor in the mid-1990s disproved this hypothesis. Subsequent studies demonstrated that HIV-1 could enter Old World monkey cells, but a block that targeted the viral capsid prevented the establishment of a permanent infection (*4*).

Using a genetic screen, we identified TRIM5α as the primary block to HIV-1 replication in Old World monkey cells (*5*). The expression of rhesus monkey TRIM5α in human cells potently inhibited HIV-1 infection and prevented the accumulation of reverse transcripts. Importantly, reducing the expression of TRIM5α in rhesus monkey cells with small interfering RNA relieved the block to HIV-1.

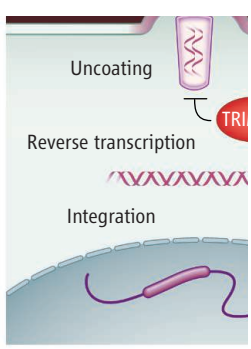
We initially hypothesized that TRIM5α functioned as a cofactor necessary for capsid uncoating. However, subsequent findings argued against this hypothesis. First, knocking down human TRIM5α showed no effects on HIV-1 replication in human cells. Second, rodent cells, which do not express TRIM5α, supported HIV-1 infection if engineered to express an appropriate receptor. Finally, human TRIM5α does not associate with the HIV-1 capsid in biochemical assays. Thus, TRIM5α appeared to have evolved primarily as an inhibitory factor aimed at thwarting viral replication, rather than a host factor co-opted by HIV-1 to promote infection.

Further evidence that TRIM5α functions primarily as a modulator of innate immunity against retroviruses emerged from comparing the sequences of *TRIM5α* orthologs from different primate species. We found dramatic length variation, and an unusually high ratio of nonsynonymous to synonymous changes in the C-terminal domain of TRIM5α orthologs (*6*) suggesting that TRIM5α has been subjected to strong positive selective pressure during primate evolution. Furthermore, episodic changes in the TRIM5α C-terminal domain coincide with periods of retroviral epidemics (*6*). Indeed, a recent report suggests that selective changes occurred in the TRIM5α lineage during acquisition of resistance to an ancient retrovirus. These changes may have had the unfortunate consequence of attenuating TRIM5α potency against HIV-1 (*7*).

Does TRIM5α have the ability to block infection by other retroviruses? We found that TRIM5α from various Old World monkey species conferred potent resistance to HIV-1, but not SIV (*5*). New World monkey TRIM5α proteins, in contrast, blocked SIV but not HIV-1 infection. Human TRIM5α inhibited N-MLV and EIAV replication (*8, 9*). Thus, the variation among *TRIM5* orthologs accounts for the observed patterns of post-entry blocks to retroviral replication among primate species.

To determine why Old World monkey TRIM5α, but not human TRIM5α, potently blocks HIV-1, we systematically altered the human sequence to more closely resemble the monkey sequence. Remarkably, we found that a single amino acid determines the antiviral potency of human TRIM5α (*10*). If a positively charged arginine residue in the C-terminal domain of human TRIM5α is either deleted or replaced with an uncharged amino acid, human cells gain the ability to inhibit HIV-1 infection (*11*). Perhaps some humans have already acquired this change and are naturally resistant to HIV-1 infection.

How does TRIM5α inhibit infection? Following viral entry into the host cell, the capsid core, which encases the viral RNA, must disassemble to allow reverse transcription of the viral RNA into DNA (see the figure). Host factors that mediate capsid uncoating are presumed to exist, but have not been identified.



Blocking HIV: Rhesus Monkey TRIM5α targets the viral capsid to block HIV-1 replication.

Because early studies demonstrated that sequences within the capsid determined susceptibility to the block, we asked if TRIM5α associated with the capsid. The association of TRIM5α with the capsid was dependent on the C-terminal domain and the association was necessary for restriction (*12*). TRIM5α proteins from various Old World monkey species bound

The author is at the U.S. Global AIDS Coordinator's Office at the State Department as a Fellow of the American Association for the Advancement of Science, Washington, DC 20522, USA. E-mail: stremlauMH@state.gov

the HIV-1 capsid; however, TRIM5α variants that did not restrict HIV, such as New World monkey TRIM5α, did not associate with the capsid cores. Human TRIM5α exhibited a very weak association with the HIV-1capsid cores, explaining the lower potency of human TRIM5α in blocking HIV-1 infection (*12*).

Why does association of TRIM5α with the viral capsid inhibit infection? Previous studies of HIV-1 capsid mutants suggest that capsid disassembly may be a temporally regulated process with either too rapid or too slow disassembly compromising viral infectivity (*13*). By following the fate of viral cores in the cytosol just after viral entry, we found that TRIM5α caused capsid cores to undergo rapid, and premature, disassembly (*12*, *14*). Accelerated uncoating of the capsid correlated with the ability of TRIM5α variants from different species to restrict HIV-1, SIV, and N-MLV infections. Future studies are needed to determine how TRIM5α promotes rapid disassembly of capsid and why accelerated disassembly is detrimental to infection. Perhaps accelerated disassembly of the retroviral capsid prematurely exposes the viral RNA or viral enzymes to degradation.

The discovery of TRIM5α not only answered a long-standing question in the HIV field, it also revealed a new pathway that protects cells from retroviral infection. The human genome encodes more than 50 members of the TRIM family. Recently, TRIM25 was shown to be essential for RIG-I–mediated antiviral activity (*15*) and TRIM family members such as TRIM1, TRIM19 (PML), and TRIM22, may block other viruses (*16*).

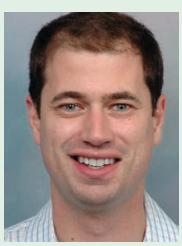
At a time when policy-makers and the public express frustration over our inability to produce an HIV vaccine, it is hoped that the discovery of TRIM5α will precipitate new ideas for how to protect human hosts from retroviral infection. Perhaps in the case of retroviruses, innate intracellular immunity mediated by factors like TRIM5α and APOBEC play a particularly crucial role. Efforts aimed at enhancing these innate immune defenses may ultimately prove to be more effective at protecting humans from HIV than vaccine strategies aimed primarily at stimulating humoral or cellular immune responses.

References

1. F. Barre-Sinoussi *et al.*, *Science* **220**, 868 (1983).
2. R. C. Gallo *et al.*, *Science* **224**, 500 (1984).
3. H. J. Alter *et al.*, *Science* **226**, 549 (1984).
4. P. D. Bieniasz, *Trends Microbiol.* **11**, 286 (2003).
5. M. Stremlau *et al.*, *Nature* **427**, 848 (2004).
6. B. Song *et al.*, *J. Virol.* **79**, 6111 (2005).
7. S. M. Kaiser, H. S. Malik, M. Emerman, *Science* **316**, 1756 (2007).
8. M. J. Perron *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11827 (2004).
9. B. Song *et al.*, *J. Virol.* **79**, 3930 (2005).
10. M. Stremlau, M. Perron, S. Welikala, J. Sodroski, *J. Virol.* **79**, 3139 (2005).
11. Y. Li, X. Li, M. Stremlau, M. Lee, J. Sodroski, *J. Virol.* **80**, 6738 (2006).
12. M. Stremlau *et al.*, *Proc. Natl. Acad. Sc.i U.S.A.* **103**, 5514 (2006).
13. B. M. Forshey, U. von Schwedler, W. I. Sundquist, C. Aiken, *J. Virol.* **76**, 5667 (2002).
14. M. J. Perron *et al.*, *J. Virol.* **81**, 2138 (2007).
15. M. U. Gack *et al.*, *Nature* **446**, 916 (2007).
16. S. Nisole, J. P. Stoye, A. Saib, *Nat. Rev. Microbiol.* **3**, 799 (2005).

10.1126/science.1152905

2007 Grand Prize Winner



Matt Stremlau, the author of the prize-winning essay and a North American regional winner, received his B.S. in chemistry from Haverford College. After graduation, he spent 1 year as a Henry Luce Fellow at the National Laboratory for Agrobiotechnology in Beijing, China, before beginning graduate studies at Harvard University. Here Dr. Stremlau investigated retroviral restriction in nonhuman primates in Dr. Joe Sodroski's laboratory. He currently works in the U.S. Global AIDS Coordinator's Office at the State Department as an American Association for the Advancement of Science Fellow. Dr. Stremlau plans to start a postdoctoral fellowship in 2008 and is interested in emerging biotechnologies relevant to the developing world.

Regional Winners

North America: Bo Huang for his essay "Molecular Accounting of a Cell." Dr. Huang was born in Chongqing, China. He graduated with honors in 2001 with a B.S. degree in chemistry from Peking University. As a graduate fellow at Stanford University, under the direction of Dr. Richard N. Zare, he worked on the development of microfluidic devices for the analysis of individual cells using single-molecule detection. Now, as a postdoctoral fellow at Harvard University, he is working with Dr. Xiaowei Zhuang on a fluorescence microscopy technique that can achieve molecular-scale resolution in biological samples.

Europe: Mikaela Rapp, for her essay "The Ins and Outs of Membrane Proteins." Dr. Rapp grew up in Stockholm, Sweden. As a Ph.D. student in the group of Dr. Gunnar von Heijne at Stockholm University, she performed a global topology analysis of the *E. coli* inner membrane proteome. Dr. Rapp defended her thesis in December 2006 and is currently learning membrane protein crystallography in the laboratory of Dr. Mika Jormakka at the Centenary Institute in Sydney, Australia. She plans to continue her scientific career as a postdoctoral fellow in the laboratory of Dr. Pär Nordlund at Karolinska Institute, Stockholm, Sweden.

Japan: Takeshi Imai for his essay "Axonal Wiring Specificity by Differential cAMP Levels of the Mouse Olfactory System." Dr. Imai was born in Tokyo in 1978 and grew up in Ina, a small southern city in Nagano, Japan. In 2001, he received a B.S. degree in biophysics and biochemistry from the University of Tokyo and remained there to pursue graduate studies in Dr. Hitoshi Sakano's laboratory, where he investigated the molecular mechanisms of the odorant receptor. He completed his Ph.D. in September 2006 but stayed on in Dr. Sakano's lab as a postdoctoral fellow.

All other countries: Manuel de la Mata for his essay "The Transcriptional Control of Alternative Splicing." Dr. de la Mata was born in Santa Rosa, Argentina. He majored in chemistry at the University of Córdoba, Argentina, and then entered a Ph.D. program at the University of Buenos Aires, where he studied the coupling of transcription with alternative splicing in the group of Dr. Alberto Kornblihtt. He defended his thesis in December 2006 and is currently a postdoctoral fellow in the Facultad de Ciencias Exactas y Naturales at the University of Buenos Aires.

For the full text of essays by the regional winners and for information about applying for next year's awards, see *Science* Online at www.sciencemag.org/feature/data/prizes/ge/index.dtl.

Optical Quantum Computing

Jeremy L. O'Brien

In 2001, all-optical quantum computing became feasible with the discovery that scalable quantum computing is possible using only single-photon sources, linear optical elements, and single-photon detectors. Although it was in principle scalable, the massive resource overhead made the scheme practically daunting. However, several simplifications were followed by proof-of-principle demonstrations, and recent approaches based on cluster states or error encoding have dramatically reduced this worrying resource overhead, making an all-optical architecture a serious contender for the ultimate goal of a large-scale quantum computer. Key challenges will be the realization of high-efficiency sources of indistinguishable single photons, low-loss, scalable optical circuits, high-efficiency single-photon detectors, and low-loss interfacing of these components.

Over the last few decades quantum information science has emerged to consider what additional power and functionality can be realized in the encoding, transmission, and processing of information by specifically harnessing quantum mechanical effects (*1*). Anticipated technologies include quantum key distribution (*2*), which offers perfectly secure communication; quantum metrology (*3*), which allows more precise measurements than could ever be achieved without quantum mechanics; and quantum lithography (*4*), which could enable fabrication of devices with features much smaller than the wavelength of light. Perhaps the most startling and powerful future quantum technology is a quantum computer, which promises exponentially faster computation for particular tasks (*1*, *5*).

The quest to develop a quantum computer will require formidable technical mastery of the fabrication of devices at the nano and possibly atomic scale, and precision control of their quantum mechanical states. The task is also daunting owing to the inherent fragility of quantum states and the fact that quantum entanglement, and its role in a quantum computer, is not yet fully understood. As we engineer devices that exploit quantum mechanical effects, we will gain an unprecedented control over the fundamental workings of nature as well as a deeper understanding of them.

The requirements for realizing a quantum computer are confounding: scalable physical qubits—two-state quantum systems—that can be well isolated from the environment but also initialized, measured, and controllably interacted to implement a universal set of quantum logic gates (*6*). However, a number of physical implementations are being pursued, including nuclear magnetic resonance, ion, atom, cavity quantum electrodynamics, solid state, and superconducting systems (*7*). Over the past few years, single particles of light—photons—have emerged as a leading approach.

Centre for Quantum Photonics, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, BS8 1UB, UK.

E-mail: Jeremy.OBrien@bristol.ac.uk

Single Photons as Qubits

Single photons are largely free of the noise, or decoherence, that plagues other systems; can be easily manipulated to realize one-qubit logic gates; and enable encoding in any of several degrees of freedom, for example, polarization, time bin, or path. Figure 1A shows how a qubit can be encoded in the polarization of a single photon. An arbitrary state of a single qubit $\alpha|H\rangle + \beta|V\rangle$ ($|\alpha|^2 + |\beta|^2 = 1$) can be represented on the Poincaré (or Bloch) sphere (Fig. 1B). One-qubit logic gates are straightforward, using birefringent wave plates (Fig. 1C), and converting between polarization and path encoding can be easily achieved using a polarizing beam splitter (PBS) (Fig. 1D), where $|0\rangle$ or $|1\rangle$ now represents a photon in the upper or lower path, respectively.

A major difficulty for optical quantum computing is in realizing the entangling logic gates required for universal quantum computation. The canonical example is the controlled NOT gate (CNOT), which flips the state of a target (T) qubit conditional on a control (C) qubit being in the logical state "1." Figure 2A shows why this operation is difficult. The two paths used to encode the target qubit are mixed at a 50% reflecting beam splitter (BS) (or half-silvered mirror), which performs the Hadamard operation (Fig. 1C). If the phase shift is not applied, the second Hadamard (BS) undoes the first, returning the target qubit to exactly the same state it started in (this is an example of "classical" wave interference). If, however, a (π) phase shift is applied, that is, $|0\rangle + |1\rangle \leftrightarrow |0\rangle - |1\rangle$, the target qubit undergoes a bit-flip or NOT operation. A CNOT must implement this phase shift only if the control photon is in the "1" path. No known or foreseen material has an optical nonlinearity strong enough to implement this conditional phase shift [although tremendous progress has been made with single atoms in high-finesse optical cavities (*8*–*10*)].

In 2001, a major breakthrough showed that scalable quantum computing is possible using

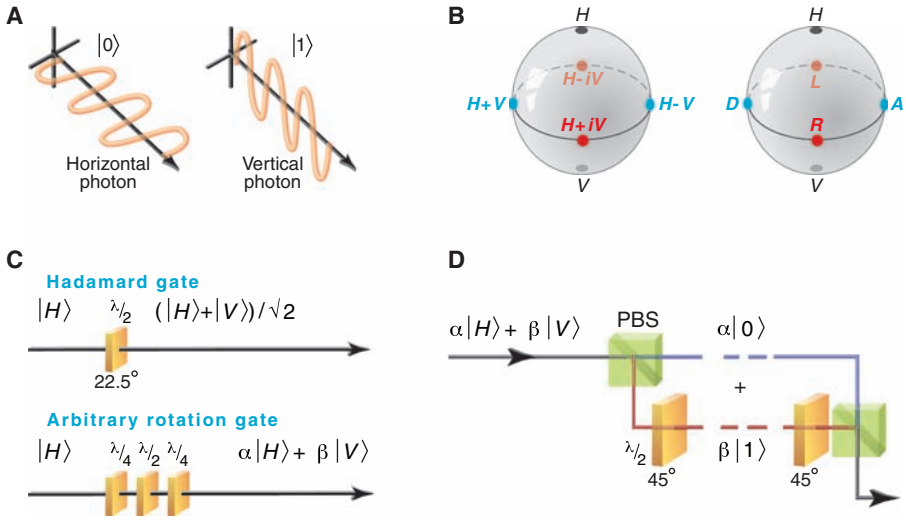


Fig. 1. Single-photon qubits. (**A**) A horizontal (H) photon represents a logical "0" and a vertical (V) photon represents a logical "1": $|0\rangle \equiv |H\rangle$; $|1\rangle \equiv |V\rangle$. (**B**) An arbitrary state can be plotted on the Bloch (or Poincaré) sphere. Examples of diagonal ($|D\rangle \equiv |0\rangle+|1\rangle$), antidiagonal ($|A\rangle \equiv |0\rangle-|1\rangle$), right circular ($|R\rangle \equiv |0\rangle+i|1\rangle$), and left circular ($|L\rangle \equiv |0\rangle-i|1\rangle$) are shown. (**C**) Single-qubit gates are easily realized using birefringent wave plates that retard one polarization by a fraction of a wavelength λ relative to an orthogonal polarization, causing a rotation of the state on the Bloch sphere, with the axis of rotation determined by the orientation of the wave plate. For example, a Hadamard (H) gate (defined by its operation on the logical states: $|0\rangle \rightarrow |0\rangle+|1\rangle$; $|1\rangle \rightarrow |0\rangle-|1\rangle$ or $|H\rangle \rightarrow |D\rangle$; $|V\rangle \rightarrow |A\rangle$) causes a π rotation about an axis running through the midpoint between the $|H\rangle$ and $|D\rangle$ states and can be realized by a $\lambda/2$ wave plate oriented at 22.5°. An arbitrary rotation requires a $\lambda/4$-$\lambda/2$-$\lambda/4$ sequence. (**D**) Converting between polarization and path encoding requires a PBS, which transmits *H* and reflects *V*, and a $\lambda/2$ wave plate oriented at 45°, which transforms $|V\rangle \leftrightarrow |H\rangle$.

only single-photon sources and detectors and simple (linear) optical circuits consisting of BSs (*11*). This is a truly remarkable discovery, because the argument above suggests that a strong optical nonlinearity is required to realize the most basic logic element.

Linear Optical Quantum Computing

A diagram of a nondeterministic (probabilistic with success signal) CNOT is shown in Fig. 3C. The control and target qubits (encoded in polarization, say), together with two auxiliary photons, enter an optical network of BSs, where the four photons' paths are combined. At the output of this network, the control and target photons emerge, having had the CNOT logic operation applied to their state, conditional on a single photon being detected at both detectors. This detection event occurs with probability $P < 1$ (1/16 in the original scheme); the rest of the time ($P = 15/16$), another detection pattern is recorded (none, only one, two photons at one detector, and so on) and the CNOT logic is not applied. In fact, a single photon may not even emerge from the control and/or target outputs in these cases.

A nondeterministic CNOT is of little use for quantum computing because the probability that a computation succeeds decreases exponentially with the number of CNOTs. Fortunately, the success probability of the nondeterministic CNOT can be boosted by harnessing quantum teleportation (*12*), a process whereby the unknown state of a qubit can be transferred to another qubit. The idea is to teleport a nondeterministic gate that has already worked onto the control and target qubits (*13*) (Fig. 3). Quantum teleportation has been realized with single photons (*14*).

An important omission from the above discussion is that, because the Bell measurements required for teleportation (Fig. 3B) measure maximally entangled states, they require a similar optical nonlinearity as a CNOT (although the photons can be destroyed in the measurement) and therefore fail some of the time. When they fail, they measure the state of the control and target photons in the $\{|0\rangle,|1\rangle\}$ basis. The final component is an encoding against this "measurement error": A single logical qubit is encoded in several physical qubits such that if one of the physical qubits is measured, the original logical qubit can still be recovered. These encoded states are entangled and therefore require entangling gates to realize them. However, by using more and more photons, a CNOT with a probability of success approaching 1 can be realized (*11*).

Reducing the Resource Overhead

These developments were expanded upon (*15*–*19*) and soon followed by several proof-of-principal experimental demonstrations of CNOTs (*20*–*23*) and encoding against measurement error (*24*, *25*). Despite this great progress, optical quantum computing was still widely regarded as impractical owing to the large resource overhead required to realize a near-deterministic CNOT: >10,000 pairs of entangled photons to achieve a success probability of >95%. The reason that all-optical quantum computing is today a promising route to practical quantum computing is due to new schemes that dramatically reduce this worrying resource overhead.

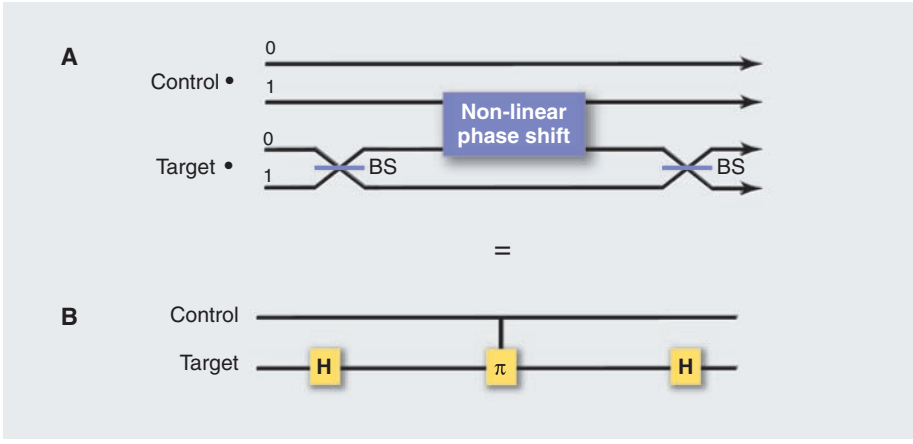


Fig. 2. An optical CNOT gate. (**A**) Schematic of a possible realization of an optical CNOT gate. (**B**) In the notation of quantum circuits, the BSs each implement a Hadamard (H) gate.

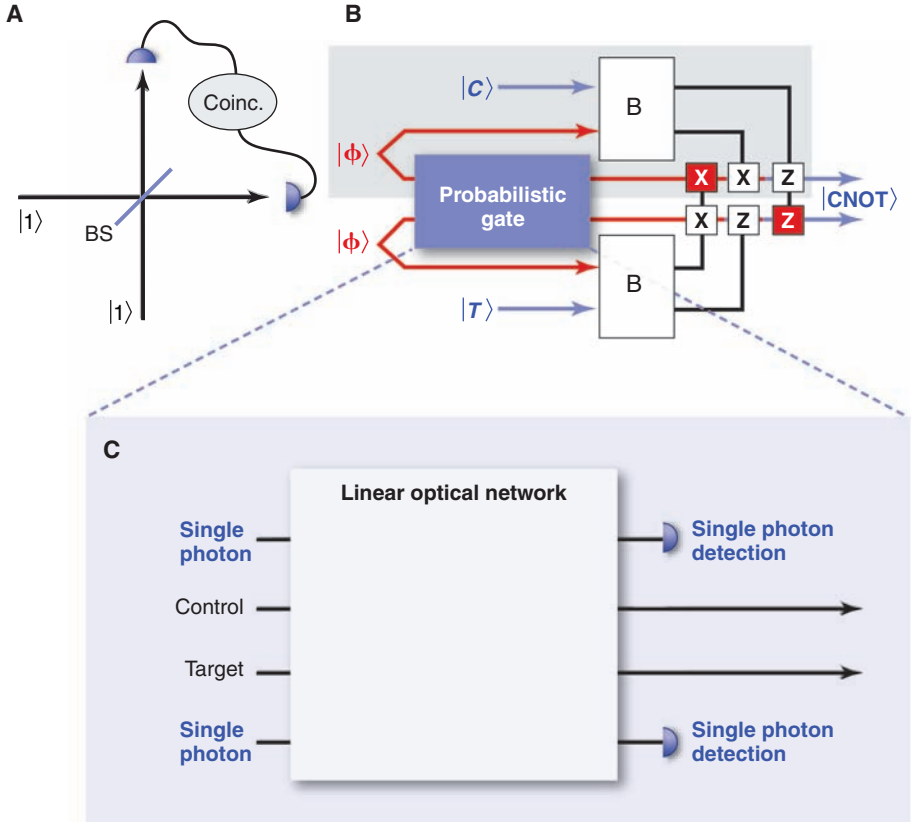


Fig. 3. An optical CNOT gate by teleportation. (**A**) Quantum interference of two photons at a BS. (**B**) Teleportation of a CNOT. Ignoring the Probabilistic Gate, a qubit in an unknown state $|C\rangle$ and one of two photons prepared in a maximally entangled state $|\Phi\rangle$ are subjected to a Bell measurement (B). This measurement leaves the third qubit in the state $|C\rangle$, or the bit (X), and/or phase (Z) flipped version of $|C\rangle$, depending on which of the four maximally entangled states is measured. An unwanted X and/or Z flip can be trivially corrected by applying a second X and/or Z as required. Still ignoring the Probabilistic Gate, the unknown input state of the control and target qubits can both be teleported and the CNOT performed on the output qubits. This seems like a lot of extra work for no gain; however, performing the CNOT before the (possible) X and Z flip has the tremendous advantage that we could repeatedly attempt the CNOT on the two halves of two entangled states, and only when the gate works would we proceed with teleportation. In this way, the control and target qubits are preserved until the gate works (on average, 32 entangled states will be consumed) and we can implement the CNOT deterministically. In quantum mechanics, the order in which operations are performed is important; performing the CNOT earlier means we must add the X and Z flips indicated in red (*13*). (**C**) A diagram of a measurement-assisted nondeterministic CNOT gate.

Quantum computations (regardless of physical realization) are typically formulated using the quantum circuit model (e.g., Fig. 2B), a generalization of the circuit model for Boolean logic: Qubits are represented by wires propagating in time from left to right, subjected to a sequence of quantum logic gates, and finally measured (*1*). In 2001 a remarkable alternative was proposed in which the computation starts with a particular massively entangled state of many qubits (a cluster state) and the computation proceeds by a sequence of single-qubit measurements, from left to right, that ultimately leave the rightmost column of qubits in the answer state (*26*) (Fig. 4).

In 2004, it was recognized that the cluster approach offered tremendous advantages for optical realizations (*27*) [see also (*28*)]. Because preparation of the cluster state can be probabilistic, nondeterministic CNOTs are suitable for making it, removing much of the massive overhead that arises from the error encoding used to make near-deterministic CNOTs. It turns out that a similar advantage can also be gained in the circuit formulation of optical quantum computing by using more sophisticated error-encoding techniques (*29*). These, and other techniques that dispense with CNOT gates entirely (*30*), reduce the resources required by 3 to 4 orders of magnitude, making an all-optical approach far more attractive. There have already been experimental proof-of-principle demonstrations of these new schemes [e.g., (*31*–*34*)].

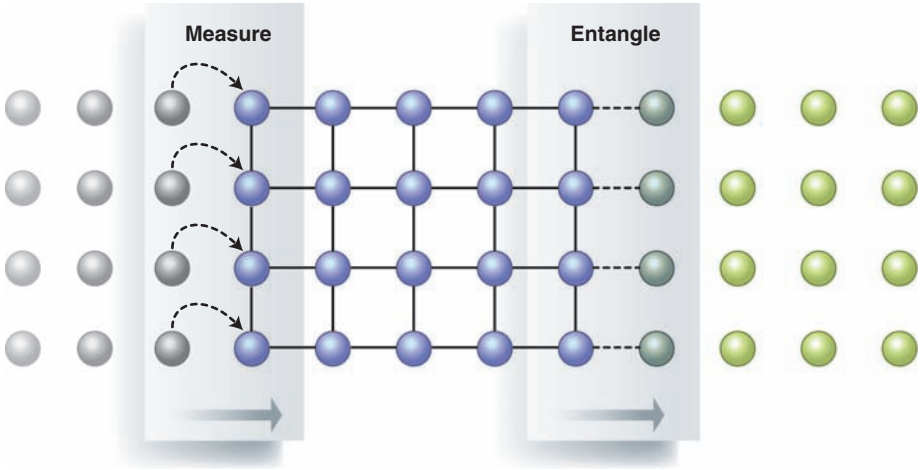


Fig. 4. Cluster state quantum computing. For photons, it is practical to start measuring the qubits while the cluster is still being grown. The blue qubits are in a cluster state, where the bonds between them represent entanglement. The green qubits are being added to the cluster, whereas the gray qubits have been measured and are no longer entangled. The measurement outcome determines the basis for the measurement on the next qubit.

Fault Tolerance

The final, and arguably most important, consideration (for all physical realizations) is fault tolerance (*1*). In contrast to conventional computers, quantum computers will be very susceptible to noise, which must be encoded against (in addition to the encoding described above). The threshold theorem says that if the noise is below some threshold, an arbitrarily long quantum computation can be realized. One of the most encouraging results for all approaches to quantum computing was the high threshold of 1% recently reported by Knill (*35*). Because cluster state approaches do not conform to the standard model, the threshold theorem does not apply; fortunately, analogous thresholds have been shown to exist (*36*). Recent results give cause for optimism: They show that if the product of source and detector efficiency is >2/3, then optical quantum computing is possible, provided all other components operate perfectly (*37*) (photon loss can in some cases be incorporated into source or detector efficiency). More complete treatments that consider more sources of noise give thresholds of 10^{-3} to 10^{-4} (*38*). The true number will likely lie somewhere in between.

Sources, Detectors, and Circuits

There are very stringent requirements for single-photon sources for optical quantum computing. In a general linear optical network (e.g., Fig. 3C) there are places where photons arrive at both inputs to a BS where quantum interference of two (or more) photons can occur. An example is shown in Fig. 3A where a photon enters each input of a 50% reflective BS. The probability of detecting a single photon at each output is given by the square of the sum of the probability amplitude for both photons to be transmitted and that for both photons to be reflected: $P = |r.r + t.t|^2$. Because a phase shift occurs on reflection $r.r = -t.t$ and so $P = 0$, in contrast to our (classical) expectation: $P = 1/2$ (*39*). For quantum interference to occur, the two photons must be indistinguishable from one another in all degrees of freedom.

To date, small-scale tests of optical quantum computing have relied on indistinguishable pairs of photons generated by a strong laser pulse in a nonlinear crystal. Unfortunately, this process is spontaneous and not readily scalable (*40*). Solid-state sources of single photons hold the promise of ready integration, and quantum interference between subsequent photons emitted from a semiconductor quantum dot has been observed (*41*). However, an optical quantum computer will require quantum interference between photons emitted from independent sources. This has very recently been achieved for a pair of trapped atoms (*42*) and ions (*43*), a tremendous advance that bodes well for optical quantum computing. Impurities in diamond may offer the best of both worlds—a solid-state host and atom-like energy levels—and have emerged as very promising candidates (*40*).

It is actually the inherent nonlinearity of photon measurement, combined with quantum interference of photons, that makes linear optical quantum computing possible. Single-photon counting modules are commercially available and have been used for almost all demonstrations to date; however, they cannot distinguish between one or more photons and have a limited efficiency (~70%). Higher efficiency will be required for scalable optical quantum computing, whereas photon number resolution will be desirable. Ongoing work indicates that such high-performance detectors will become available, with superconductor-based devices holding great promise (*44*).

Finally, almost all demonstrations of linear optical logic circuits have relied on large-scale BSs and mirrors, with photons propagating in air; improved performance, miniaturization, and scalability will likely require low-loss microscopic optical waveguide circuits. A promising approach is integrated optics, an analog of electrical integrated circuits, which has been developed by the photonics industry. Outstanding challenges are to realize quantum interference in these devices and to integrate them with single-photon sources and detectors.

Nonlinear and Hybrid Approaches

Recently, attention has been given to the idea of combining linear optics with optical nonlinearities that would not allow a CNOT gate to be realized in the manner suggested in Fig. 2 but would nevertheless offer considerable advantages. One way is to use a two-photon absorber to implement the quantum Zeno effect, whereby repeated measurement inhibits the emission of two photons into one of the outputs of a CNOT gate (*45*)—the failure mode of the linear optical CNOT gate proposed in (*16*). Another way is to use a strong optical nonlinearity that is significantly weaker than that required in Fig. 2: Single photons are made to interact with one another by means of a bright laser pulse and the nonlinear medium (*46*). Finally, recent developments suggest that a hybrid approach may have many advantages (*47*–*50*): Because single-photon sources are inherently quantum mechanical, it is prom-

ising to consider storing quantum information in the sources themselves; already, spins associated with impurities in diamond have shown great promise in this direction (*51*). Such systems are particularly suited to the small-scale quantum processors that will be required in the nodes and quantum repeaters of quantum communication networks (*52*).

Future Prospects

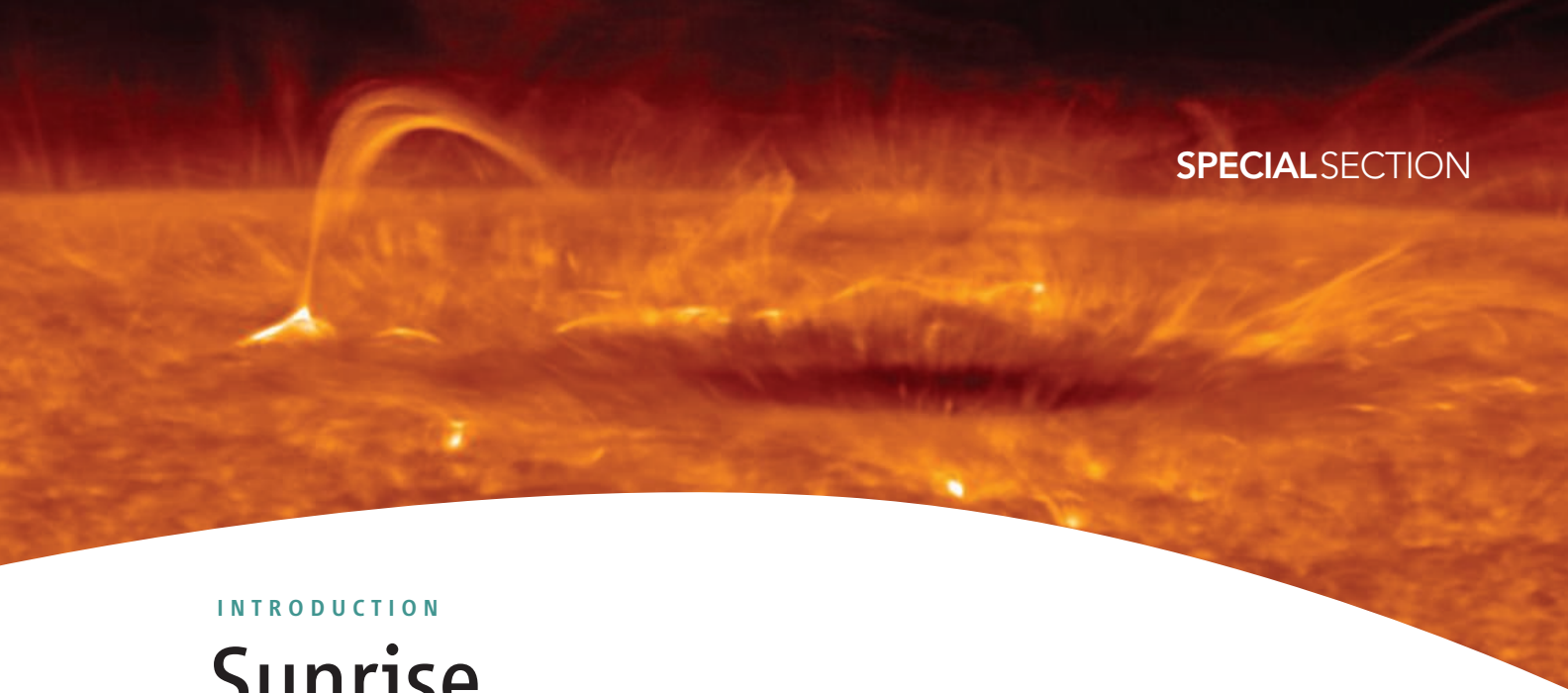
Despite great progress, much work remains to be done if a large-scale optical quantum computer is to be realized. It is not yet known whether the circuit or cluster model (or some other approach) is most promising. Indeed, a combination of these approaches has been described in which error encoding is achieved using cluster techniques but the computation proceeds through conventional CNOT gates (*53*). Further, the role of nonlinear optics approaches in any future optical quantum computer will depend on their efficacy and practicality. The majority of experimental demonstrations to date have relied on nonscalable single-photon sources, large-scale optical elements, and modest-efficiency single-photon detectors. Scaling to useful devices will require high-efficiency single-photon sources and detectors that are efficiently coupled to low-loss microscopic optical waveguide circuits [optical memories may not be required (*54*)].

References and Notes

1. M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
2. N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, *Rev. Mod. Phys.* **74**, 145 (2002).
3. V. Giovannetti, S. Lloyd, L. Maccone, *Science* **306**, 1330 (2004).
4. A. N. Boto *et al*., *Phys. Rev. Lett.* **85**, 2733 (2000).
5. D. Deutsch, *Proc. R. Soc. Lond. A* **400**, 97 (1985).
6. D. P. DiVincenzo, D. Loss, *Superlatt. Micro.* **23**, 419 (1998).
7. See the U.S. Advanced Research and Development Activity (ARDA) Quantum Computation Roadmap for current state-of-the-art; http://qist.lanl.gov/qcomp_map.shtml.
8. Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, H. J. Kimble, *Phys. Rev. Lett.* **75**, 4710 (1995).
9. T. Aoki *et al*., *Nature* **443**, 671 (2006).
10. M. Hijlkema *et al*., *Nature Phys.* **3**, 253 (2007).
11. E. Knill, R. Laflamme, G. J. Milburn, *Nature* **409**, 46 (2001).
12. C. H. Bennett *et al*., *Phys. Rev. Lett.* **70**, 1895 (1993).
13. D. Gottesman, I. L. Chuang, *Nature* **402**, 390 (1999).
14. D. Bouwmeester *et al*., *Nature* **390**, 575 (1997).
15. M. Koashi, T. Yamamoto, N. Imoto, *Phys. Rev. A* **63**, 030301 (2001).
16. T. B. Pittman, B. C. Jacobs, J. D. Franson, *Phys. Rev. A* **64**, 062311 (2001).
17. T. C. Ralph, A. G. White, W. J. Munro, G. J. Milburn, *Phys. Rev. A* **65**, 012314 (2001).
18. T. C. Ralph, N. K. Langford, T. B. Bell, A. G. White, *Phys. Rev. A* **65**, 062324 (2001).
19. H. F. Hofmann, S. Takeuchi, *Phys. Rev. A* **66**, 024308 (2001).
20. T. B. Pittman, M. J. Fitch, B. C. Jacobs, J. D. Franson, *Phys. Rev. A* **68**, 032316 (2003).
21. J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, D. Branning, *Nature* **426**, 264 (2003).
22. J. L. O'Brien *et al*., *Phys. Rev. Lett.* **93**, 080502 (2004).
23. S. Gasparoni, J.-W. Pan, P. Walther, T. Rudolph, A. Zeilinger, *Phys. Rev. Lett.* **93**, 020504 (2004).
24. J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, *Phys. Rev. A* **71**, 060303 (2005).
25. T. B. Pittman, B. C. Jacobs, J. D. Franson, *Phys. Rev. A* **71**, 052332 (2005).
26. R. Raussendorf, H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188 (2001).
27. M. A. Nielsen, *Phys. Rev. Lett.* **93**, 040503 (2004).
28. N. Yoran, B. Reznik, *Phys. Rev. Lett.* **91**, 037903 (2003).
29. T. C. Ralph, A. J. F. Hayes, A. Gilchrist, *Phys. Rev. Lett.* **95**, 100501 (2005).
30. D. E. Browne, T. Rudolph, *Phys. Rev. Lett.* **95**, 010501 (2005).
31. P. Walther *et al*., *Nature* **434**, 169 (2005).
32. N. Kiesel *et al*., *Phys. Rev. Lett.* **95**, 210502 (2005).
33. R. Prevedel *et al*., *Nature* **445**, 65 (2007).
34. C.-Y. Lu *et al*., *Nat. Phys.* **3**, 91 (2007).
35. E. Knill, *Nature* **434**, 39 (2005).
36. M. A. Nielsen, C. M. Dawson, *Phys. Rev. A* **71**, 042323 (2005).
37. M. Varnava, D. Browne, T. Rudolph, *arXiv:quant-ph/0702044* (2007).
38. C. M. Dawson, H. L. Haselgrove, M. A. Nielsen, *Phys. Rev. A* **73**, 052306 (2006).
39. Quantum mechanics tells us that if a particular event can happen in two or more indistinguishable ways, we must first sum the probability amplitudes before squaring to obtain a probability. Because these amplitudes can be negative, it is possible for events that we would intuitively expect to be possible to have zero probability. If the BS is more or less reflective than 50%, these two amplitudes no longer have the same magnitude but are still opposite in sign, meaning that the probability is nonzero but less than one would naively expect.
40. See special issue: Focus on Single Photons on Demand, Eds., P. Grangier, B. Sanders, and J. Vuckovic, *New J. Phys.* **6** (2004).
41. C. Santori, D. Fattal, J. Vuckovic, G. S. Solomon, Y. Yamamoto. *Nature* **419**, 594 (2002).
42. J. Beugnon *et al*., *Nature* **440**, 779 (2007).
43. P. Maunz *et al*., *Nat. Phys.* **3**, 538 (2007).
44. See special issue: Single-photon: detectors, applications, and measurement methods, Eds., A. Migdal and J. Dowling, *J. Mod. Opt.* **51** (2004).
45. J. D. Franson, B. C. Jacobs, T. B. Pittman, *Phys. Rev. A* **70**, 062302 (2004).
46. S. D. Barrett *et al*., *Phys. Rev. A* **71**, 060302 (2005).
47. D. E. Browne, M. B. Plenio, S. F. Huelga, *Phys. Rev. Lett.* **91**, 067901 (2003).
48. S. C. Benjamin, J. Eisert, T. M. Stace, *New J. Phys.* **7**, 194 (2005).
49. Y.-L. Lim, S. D. Barrett, A. Beige, P. Kok, L. C. Kwek, *Phys. Rev. A* **73**, 012304 (2006).
50. S. J. Devitt *et al*., *Phys. Rev. A* **76**, 052312 (2007).
51. F. Jelezko, J. Wrachtrup, *J. Phys. Condens. Matt.* **16**, 1089 (2004).
52. L. Childress, J. M. Taylor, A. Sørensen, M. D. Lukin, *Phys. Rev. Lett.* **96**, 070504 (2006).
53. A. Gilchrist, A. J. F. Hayes, T. C. Ralph, *Phys. Rev. A* **75**, 052328 (2007).
54. K. Kieling, T. Rudolph, J. Eisert, arXiv:quant-ph/0611140 (2006).

55. I am indebted to all my collaborators past and present who have helped in my understanding of this subject. Optical quantum computing research at Bristol is supported by the U.K. Engineering and Physical Sciences Research Council, the U.K. Quantum Information Processing Interdisciplinary Collaboration, the U.S. Disruptive Technologies Office, the E.U. Integrated Project Qubit Applications, the Leverhulme Trust, and the Daiwa Anglo-Japanese Foundation.

10.1126/science.1142892



INTRODUCTION

Sunrise

GIVEN THE IMPORTANCE OF THE SUN TO SUSTAINING LIFE ON EARTH FROM season to season, it is not surprising that some of the first scientific observations were of the Sun or that the sunspot record is one of our longest scientific records. Nor is it surprising that exploration of the Sun has continued vigorously, more recently through a wide range of approaches using computer simulations, neutrino detectors, and spacecraft sampling of the solar wind, and of course a wide variety of telescope observations. Yet many of the Sun's mechanisms remain a mystery. We know that nuclear reactions in its core fuel its energy and magnetism and the emission of most of its light from its nearly 6000-kelvin surface, though many questions remain about the details of all of these processes. Moving out from the Sun's surface, things get particularly interesting. The solar corona, which is orders of magnitude less dense than the surface region, is heated to near 3,000,000 kelvin. The mechanism that transfers energy from the solar surface to heat the corona has remained a mystery. This heating, as well as the corona's intense magnetic fields and loops, stretch the ionized gas of the corona out to millions of kilometers and produce solar flares and the solar wind.

Hinode (or "sunrise") is a solar space telescope mission spearheaded by the Japan Aerospace Exploration Agency, in collaboration with partners in the United States and United Kingdom, aimed at addressing many of the questions associated with the Sun's corona. It was launched in September 2006 and began observing the Sun about 1 month later from its orbit around Earth. The papers in this special issue, as highlighted in a Perspective by Erdélyi and Fedun (p. 1572), present some of the first results from Hinode, including important clues to the mechanisms heating the corona and the acceleration of the solar wind. Hinode has spectrometers that can view the Sun in optical wavelengths as well as in extreme-ultraviolet and x-ray wavelengths that can image structures and magnetic fields within the high-energy solar plasma. It provides particularly high resolution in time and space, allowing the dynamics of small wavelike structures to be captured as movies, and revealing features, such as fine-scale jets related to magnetic fields in the corona, that have not been seen previously. New information is also provided about how magnetic lines may cross and energetically reconnect, causing solar flares. Observations of one type of wave, known as an Alfvén wave and involving a traveling oscillation of the magnetic field and plasma along a magnetic loop, may resolve the corona heating problem and explain the acceleration of the solar wind. Much more is expected to come from Hinode, yet it has already provided a new view of the Sun well-fitting its name.

–BROOKS HANSON AND DAVID VOSS

Science

CREDIT: Y. KATSUKAWA

Hinode

CONTENTS

Perspective

Reports

PERSPECTIVE

Are There Alfvén Waves in the Solar Atmosphere?

R. Erdélyi* and V. Fedun

The Sun's outer coronal layer exists at a temperature of millions of kelvins, much hotter than the solar surface we observe. How this high temperature is maintained and what energy sources are involved continue to puzzle and fascinate solar researchers. Recently, the Hinode spacecraft was launched to observe and measure the plasma properties of the Sun's outer layers. The data collected by Hinode reveal much about the role of magnetic field interactions and how plasma waves might transport energy to the corona. These results open a new era in high-resolution observation of the Sun.

The solar coronal plasma, an ionized gas that is mainly confined by the magnetic field of the Sun, is very hot—several million kelvin. Surprisingly, the temperature of the hot corona is about three orders of magnitude higher than the temperature of the visible solar surface. The heating processes that generate and sustain the hot corona have so far defied quantitative understanding, despite efforts spanning more than half a century. There are at least three fundamental questions that must be resolved in order to understand the heating of the solar atmosphere: Where is the energy generated? How does the generated energy propagate from the energy reservoir to the solar corona? How does the transported energy dissipate efficiently in the solar corona to maintain its multimillion-kelvin temperature?

A strongly related and also unsolved question is the generation and acceleration of the solar wind that often violently interacts with the magnetosphere and upper atmosphere of Earth, forming a protecting shield in space for humankind. Hinode (which means "sunrise" in Japanese, launched by the Japanese Space Agency on 22 September 2006) joins a number of other outstanding high-resolution contemporary space missions, such as Yohkoh, the Solar and Heliospheric Observatory (SOHO), the Transition Region and Coronal Explorer (TRACE), the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI), and the Solar Terrestrial Relations Observatory (STEREO), all of which have been used to study the heating of the solar corona, the acceleration of solar wind, and solar plasma particles. Hinode, with its superb high-resolution instrumental suite including the Solar Optical Telescope (SOT) (*1*), the Extreme Ultraviolet Imaging Spectrometer (EIS) (*2*), and the X-Ray Telescope (XRT) (*3*), has now revealed some of the secrets of the coronal heating problem, the acceleration of solar wind, and the generation of a wide range of energetic plasma jets.

It is clear now that the crucial ingredient to understanding the corona is the ubiquitous magnetic fields in the solar atmosphere. These fields can interact or reconnect with each other, resulting in energetic events such as dynamic jets. Some of these jets may contribute to coronal heating. Among the many theories, perhaps the most compelling one involves Alfvén waves, predicted theoretically by Hannes Alfvén, for which he received a Nobel Prize in 1970. However, no ground-based or space missions could unambiguously prove the existence of these special magnetic waves. Has Hinode broken the observational barriers facing solar physicists and astrophysicists, and has it indeed detected Alfvén waves in the solar atmosphere?

Observations of the solar corona go back at least three millennia. One of the earliest records of the corona was made by Babylonian astronomers, who reported during a solar eclipse in 1063 BCE that "the day was turned to night, and fire in the midst of the heaven." The discovery of spectroscopy in the late 19th century revealed the existence of the coronal "green line" at 5303 Å that puzzled astrophysicists for 50 years. The observed wavelength of this spectral line did not match any known elements and was thought to represent a new element (coronium). The solution of this solar mystery came from Edlén in 1939, who showed that the coronium line is emitted by highly ionized iron at a temperature well over 10^6 K.

Modern satellite observations from Skylab in the 1970s through the Solar Maximum Mission (SMM), Yohkoh, and in present times SOHO, TRACE, RHESSI, and STEREO have revealed the solar atmosphere with unprecedented spatial and temporal resolution covering wavelengths from ultraviolet through soft and hard x-rays to even gamma rays. Both spectroscopic and imaging instruments have contributed to the discovery of the large number of various fine-scale structures in the solar atmospheric zoo, from small-scale x-ray bright points to large solar coronal loops, all showing a common feature: a very sharp temperature rise. Maintaining this high temperature requires some sort of additional energy, because without it the corona would cool down by small-scale thermodynamic relaxation. Surprisingly, the nonthermal energy excess needed to sustain the solar corona is a small fraction of the total solar output.

It is relatively straightforward to estimate the entire energy budget needed for the solar corona: Approximately 10^{-4} of the Sun's total energy output is needed. Thus, at least in theory, it should be a fairly easy task for theoreticians to put forward various mechanisms that could divert 0.01% of the total solar output into heating the corona. The subphotospheric convection zone has more than enough kinetic energy to supply the observed temperature profile in the atmosphere. The question is not where the coronal nonthermal energy comes from, but rather how it actually gets to the corona from the convection zone and how it dissipates efficiently there. Hinode's main discoveries so far contribute to finding additional observational evidence of ubiquitous magnetic waves that may be the long-sought candidate for transporting the energy from the Sun's subsurface energy reservoir to the solar atmosphere.

There are two popular and well-accepted classes of theories about how energy can be transported to the corona. One addresses the key feature of building up considerable magnetic stress by random shaking of the magnetic field lines deeply rooted in subphotospheric regions. In this scenario, originally put forward by Parker (*4*) and numerically confirmed by Yokoyama and Shibata (*5*), energy will be released through the reconnection of neighboring magnetic field lines with locally opposite polarity vector components. A consequence of this violent release of energy will be two ejected jets at the reconnection site that are accelerated and repelled by the reconnected and highly curved magnetic field lines. Hinode has given proof, with outstanding quality, of the existence of a large number of these jets.

Shibata *et al*. (*6*) report a much higher number of chromospheric examples of anemone jets, as expected in sunspot active regions. These narrow (few hundred kilometers in diameter) collimated upside-down Y–shaped jets are the direct consequences of magnetic reconnection. Their propagation speed, predicted by theory to be the Alfvén speed, was also confirmed by SOT observations. Whether their contribution is indeed relevant to coronal heating will depend on the statistical properties, including their power-law index (which still has to be determined), of these energetic jets often called nanoflares.

Numerous small-scale and short-lived jets were also detected by Katsukawa *et al.* (*7*) in the penumbral chromosphere, whereas De Pontieu *et al.* (*8*) found them almost everywhere in the chromosphere. Jets may also occur on a much larger scale. Cirtain *et al*. (*9*) report Hinode XRT observations of giant (2 to 20 Mm wide, 100 Mm long) jets, another proof of magnetic reconnection

Solar Physics and Space Plasma Research Centre, Department of Applied Mathematics, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield S3 7RH, UK.

*To whom correspondence should be addressed. E-mail: robertus@sheffield.ac.uk

taking place at a much larger scale. Not only are these jets manifestations of magnetic reconnection, they are also key ingredients of the continuous mass loss of the Sun, called solar wind (*10*, *11*). Sakao *et al*. (*12*) also observed such massive outflows in solar active regions. However, the estimated mass flow rate does not seem to be sufficient to explain the total mass outflow in the solar wind, so the question still remains: What is the complete origin of these solar outflows?

The second class of theories for a viable coronal heating scenario involves magnetohydrodynamic (MHD) waves. In uniform plasma, there are three distinct types of MHD waves: slow and fast magnetoacoustic waves, and Alfvén waves. The first two types of waves have an acoustic character modified by the magnetic field, whereas the Alfvén waves exist purely because of the presence of a magnetic field. There are numerous examples of slow and fast MHD waves observed in the solar atmosphere [for a recent review, see Banerjee *et al*. (*13*)]; however, their energy density does not seem to be near enough to explain the coronal heating. On the other hand, among MHD wave theorists, the Alfvén waves are considered the most promising energy transporter. It was always expected that once these waves are generated, they will easily propagate along magnetic flux tubes, the building block of the solar atmosphere (Fig. 1A), or along magnetic field lines at constant magnetic surfaces (Fig. 1B).

According to Cirtain *et al*. (*9*) and De Pontieu *et al*. (*8*), signatures of Alfvén waves were observed with the use of the XRT and SOT instruments, respectively. Because these jets are increases in local density, if their length is short relative to the wavelength of Alfvénic perturbations, they can slide along magnetic field lines (see the yellow-red blob in Fig. 1B) in a plasma that has almost no resistivity, just like bobsleds would run and slide along a prebuilt

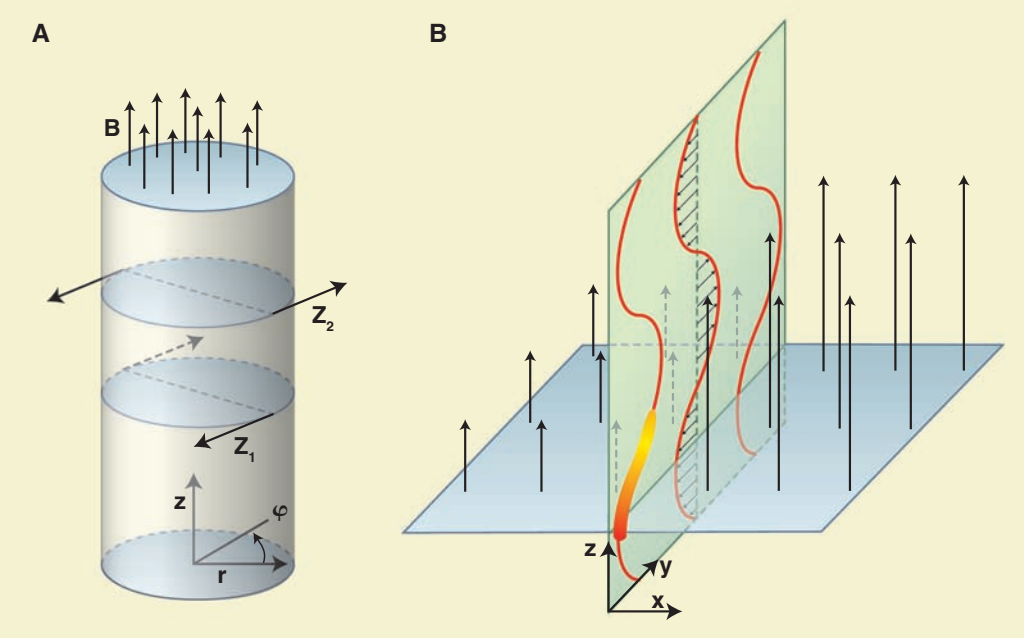


Fig. 1. (**A**) Magnetic flux tube showing two snapshots (at positions Z_1 and Z_2) of Alfvén wave perturbations propagating in the longitudinal *z* direction along field lines. At a given position, the Alfvénic perturbations are torsional oscillations (i.e., oscillations are in the φ direction, perpendicular to the background field). (**B**) Alfvén waves propagating along a magnetic discontinuity. The Alfvénic perturbations are within the magnetic surface (*yz* plane) at the discontinuity, perpendicular to the background field (*y* direction), whereas the waves themselves propagate along the field lines (*z* direction). Density enhancements [e.g., in the form of spicules as observed by De Pontieu *et al*. (*8*) or within prominences as seen by Okamoto *et al*. (*14*)] are visualized here as a yellow-red thin blob that follows the field lines. Vertical arrows indicate the magnetic field gradient increasing from left to right.

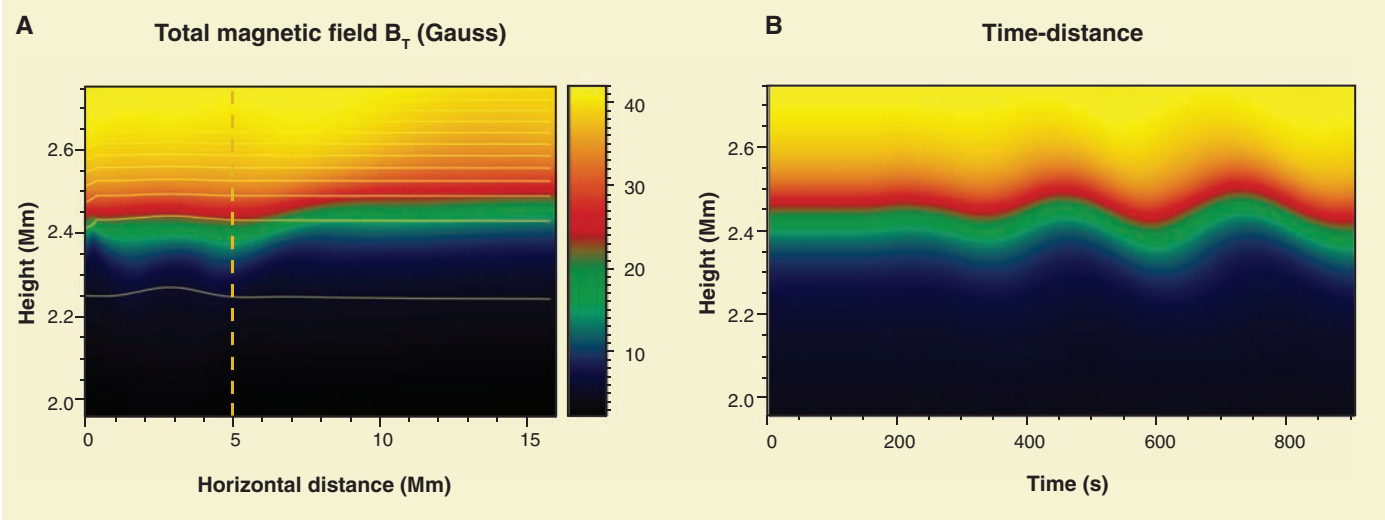


Fig. 2. (**A**) Simulation of wave perturbations along horizontal magnetic field lines (white lines in the *x* direction) embedded in a vertically stratified (*y* direction) solar atmosphere. The color bar indicates the total magnetic intensity perturbations. (**B**) Time-distance image showing the passage of the wave at an arbitrarily chosen location [vertical dashed line at $x = 5$ Mm in (A)]. Although the resemblance of the time-distance plot to (for example) the observations of waves found by Okamoto *et al*. (*14*) is striking, these waves are not Alfvénic in nature because the perturbations are not at constant magnetic surfaces (*xy* plane).

track. Perhaps most important, De Pontieu *et al.* have also estimated the energy density of the observed waves, and for the first time it has been clearly shown that the energy associated with the waves is more than enough to accelerate the solar wind and heat the quiet corona. These pioneering results have serious consequences for solar and stellar coronal heating theories.

However, these observations also raise concerns about the applicability of the classical concept of a magnetic flux tube in the apparently very dynamic solar atmosphere, where these sliding jets were captured. In a classical magnetic flux tube, propagating Alfvén waves along the tube would cause torsional oscillations (Fig. 1A). In this scenario, the only observational signature of Alfvén waves would be spectral line broadening. Hinode does not have the appropriate instrumentation to carry out line width measurements. On the other hand, if these classical flux tubes did indeed exist, then the observations of De Pontieu *et al.* (*8*) would be interpreted as kink waves (i.e., waves that displace the axis of symmetry of the flux tube like an S-shape). More detailed observations are needed, perhaps jointly with STEREO, so that a full three-dimensional picture of wave propagation would emerge.

Solar prominences, heavy and cool elongated magnetic structures supported by mainly horizontal magnetic fields in the solar corona, also support a wide variety of MHD waves. Okamoto *et al.* (*14*) report for the first time, using SOT observations, the presence of Alfvénic perturbations in a prominence. Analyzing the properties of these waves can give unprecedented insight into the fine structure with detailed diagnostic information and a clue about the magnetic field strength of a prominence by means of solar magnetoseismology. However, again we must exercise caution in any interpretations, because perturbations only in the plane of the sky were reported. Numerical simulations of a two-dimensional stratified VAL atmosphere model embedded in a horizontal magnetic field mimicking solar prominences, driven by global photospheric motions, will also result in observational signatures very similar to those reported by Okamoto *et al.* (*14*) (see, e.g., Fig. 2 and movie S1). In particular, note the similarity between the time-distance plot (Fig. 2A) of the illustrative forward modeling simulations and figure 3 of Okamoto *et al.* (*14*). However, in the case of this numerical experiment the oscillations found are clearly not Alfvénic. Again, a joint Hinode/STEREO campaign may shed light on the full nature of the observed waves.

Hinode, with its three high-resolution and high-speed instruments on board, has opened new avenues for solar observation and theory. The unprecedented detection of numerous, highly dynamic fine-scale structures mainly following the magnetic field lines will serve as an impetus for explaining the dynamics and heating of the solar atmosphere. The new insights into the details of magnetic reconnection, the origin of solar wind, the direct and unprecedented proof of existence of Alfvénic waves in a wide range of solar magnetic structures (e.g., from sunspot penumbrae to coronal holes), and their role in coronal heating will surely generate further fruitful discussions in the solar and stellar community.

References and Notes

1. T. Kosugi *et al.*, *Solar Phys.* **243**, 3 (2007).
2. J. L. Culhane *et al.*, *Solar Phys.* **243**, 19 (2007).
3. L. Golub *et al.*, *Solar Phys.* **243**, 63 (2007).
4. E. N. Parker, *Astrophys. J.* **330**, 474 (1988).
5. T. Yokoyama, K. Shibata, *Nature* **375**, 42 (1995).
6. K. Shibata *et al.*, *Science* **318**, 1591 (2007).
7. Y. Katsukawa *et al.*, *Science* **318**, 1594 (2007).
8. B. De Pontieu *et al.*, *Science* **318**, 1574 (2007).
9. J. W. Cirtain *et al.*, *Science* **318**, 1580 (2007).
10. S. Chapman, *Smithsonian Contr. Astrophys.* **2**, 1 (1957).
11. E. N. Parker, *Astrophys. J.* **128**, 664 (1958).
12. T. Sakao *et al.*, *Science* **318**, 1585 (2007).
13. D. Banerjee, R. Erdélyi, R. Oliver, E. O'Shea, *Solar Phys.* **246**, 136 (2007).
14. T. J. Okamoto *et al.*, *Science* **318**, 1577 (2007).
15. R.E. acknowledges M. Kéray for patient encouragement and is grateful to NSF Hungary (OTKA K67746).

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1572/DC1
Movie S1

10.1126/science.1153006

REPORT

Chromospheric Alfvénic Waves Strong Enough to Power the Solar Wind

B. De Pontieu,[1]* S. W. McIntosh,[2,3] M. Carlsson,[4] V. H. Hansteen,[4,1] T. D. Tarbell,[1] C. J. Schrijver,[1] A. M. Title,[1] R. A. Shine,[1] S. Tsuneta,[5] Y. Katsukawa,[5] K. Ichimoto,[5] Y. Suematsu,[5] T. Shimizu,[6] S. Nagata[7]

Alfvén waves have been invoked as a possible mechanism for the heating of the Sun's outer atmosphere, or corona, to millions of degrees and for the acceleration of the solar wind to hundreds of kilometers per second. However, Alfvén waves of sufficient strength have not been unambiguously observed in the solar atmosphere. We used images of high temporal and spatial resolution obtained with the Solar Optical Telescope onboard the Japanese Hinode satellite to reveal that the chromosphere, the region sandwiched between the solar surface and the corona, is permeated by Alfvén waves with strong amplitudes on the order of 10 to 25 kilometers per second and periods of 100 to 500 seconds. Estimates of the energy flux carried by these waves and comparisons with advanced radiative magnetohydrodynamic simulations indicate that such Alfvén waves are energetic enough to accelerate the solar wind and possibly to heat the quiet corona.

The energy source driving the acceleration of the solar wind and heating of the quiet corona remains unknown. One promising candidate is Alfvén waves, transverse magnetohydrodynamic (MHD) waves that can propagate along the magnetic field over large distances and transport magnetoconvective energy from near the surface into the outer atmosphere (*1*–*7*). Alfvén waves of sufficient strength to drive the solar wind have never been directly observed in the lower solar atmosphere. This uncertainty has cast doubts on models of a wave-driven wind, because crucial model input parameters, such as the frequency spectrum and energy flux, are unknown (*6*).

We report here direct observations of Alfvén waves that carry an energy flux (100 W m^{-2}) into the corona that is sufficient to drive the solar wind. Until now, Alfvén waves had been difficult to observe by means of remote sensing of the solar atmosphere because of the limited resolution of available instruments. The high spatial (150 km on the Sun) and temporal (5-s) resolution of the Solar Optical Telescope (SOT) (*8*) on board the recently launched Hinode satellite (*9*) has allowed us to resolve some of the dominant spatial and temporal scales of the lower atmosphere, as well as the predicted amplitudes of the Alfvén waves (*6*). As a result, our observations suffer far less from the effects of superposition of many unresolved, independently moving structures along a long line of sight in an optically thin plasma that have plagued lower-resolution instruments in the past.

We analyzed several time series of chromospheric Ca II H-line (3968 Å) images taken with a broadband filter on board the SOT (full width at half maximum, 2 Å). These movies reveal how the chromosphere is dominated by a multitude of thin (~200 km wide), dynamic, jetlike extrusions called spicules. Spicules shoot upward at speeds between 20 and 150 km/s, reaching heights between 2000 and 10,000 km [for a description of their general properties, see (*10*)]. Here, we report that many of these chromospheric spicules undergo substantial transverse displacements on the order of 500 to 1000 km during their short lifetimes of 10 to 300 s (most are less than 100 s). Some longer-lived spicules undergo a swaying or oscillatory motion in a direction perpendicular to

their own axis (Fig. 1 and movies S1 and S2), with the displacement varying sinusoidally in time (with a period of 3 min in Fig. 1 and movie S1). Spicules outline the direction of the magnetic field, because they are formed at heights where the magnetic field dominates the dynamics of the plasma (plasma $\beta < 1$, where β is the ratio of the gas pressure to the magnetic pressure). As a result, the oscillatory motion in a direction transverse to the long axis of the spicule implies the presence or passage of Alfvénic wave motions. Here, we use the term "Alfvén waves" to describe incompressible transverse MHD waves that propagate along the magnetic field in an inhomogeneous medium (*1*, *4*–*7*). The observed waves could also be interpreted as MHD kink-mode waves, should a stable waveguide exist in the chromosphere (*11*).

Our analysis of the Hinode data reveals that the Alfvénic motions are ubiquitous in the upper chromosphere. Because of the substantial line-of-sight superposition at the limb, these motions are best seen in data in which spatial scales on the order of 150 to 200 km (the typical spicule width) have been enhanced (using unsharp masking). A space-time plot along a cut parallel to the solar limb at a low height of 1000 km above the limb shows a myriad of mostly linear features (Fig. 2), which indicates that many spicules are moving transversely to their own axis at roughly constant speed. The predominance of linear motion, as opposed to full oscillatory swings, for what are in fact oscillations is not surprising if we consider the lifetime of spicules relative to the wave period. We have performed Monte Carlo simulations (*12*) that show that linear motion visually appears to dominate if the lifetime of the spicules carrying the Alfvén waves is much shorter than the wave period (fig. S2). The superposition of many independent bright features that carry Alfvénic motions with random phases leads to poor visibility of the extrema or swings in the sinusoidal motion (almost vertical in space-time plots), because many of the sinusoidal swings are superimposed on top of features that show little apparent lateral motion (e.g., because the polarization direction of the Alfvén wave is along the line of sight). Comparison of observations with simulations (Fig. 2, B and C) confirms that only the motions at high speed (which are almost linear for a sinusoidal path during the phase in between extrema of a sinusoidal motion) generate a space-time (xt) signal strong enough to catch the eye. Higher above the limb there are fewer spicules, so superposition is less of a problem. Many of these tall features also have longer lifetimes. As a result, cuts high above the limb reveal clearer swings, as

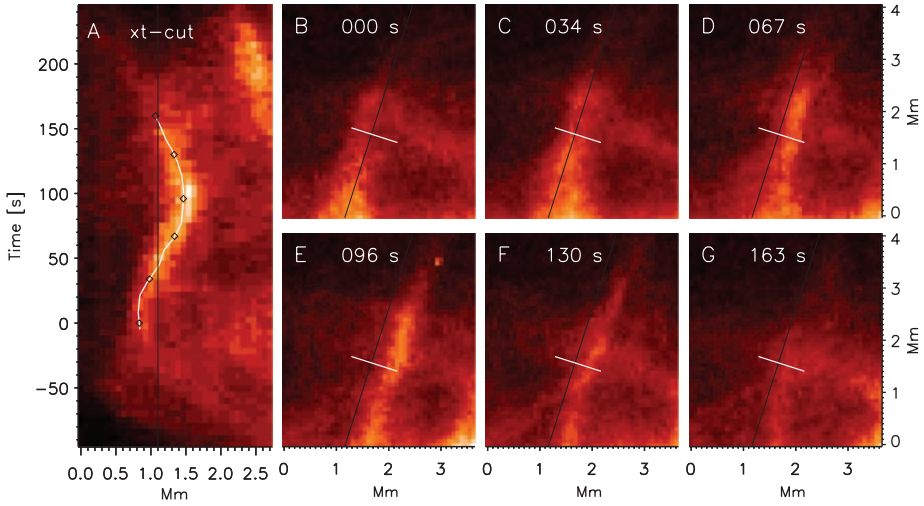


Fig. 1. Example of the transverse displacement of a spicule. (**A**) The intensity as a function of time (in seconds) along the spatial cut (in megameters) shown by the white line in (B) to (F) (space-time or xt plot). Black diamonds indicate the spicule location at the times shown in (B) to (F). The spicule sways from left at time (t) = 0 s to right at t = 96 s, and back to the starting position. These motions are compatible with the propagation along the spicule of an Alfvén wave with a large wavelength (>4 Mm). (**B** to **G**) A time series of Ca II H 3968 Å images from the Hinode SOT (movie S1). The white line shows the extent of the transverse displacement of the spicule (black guide line).

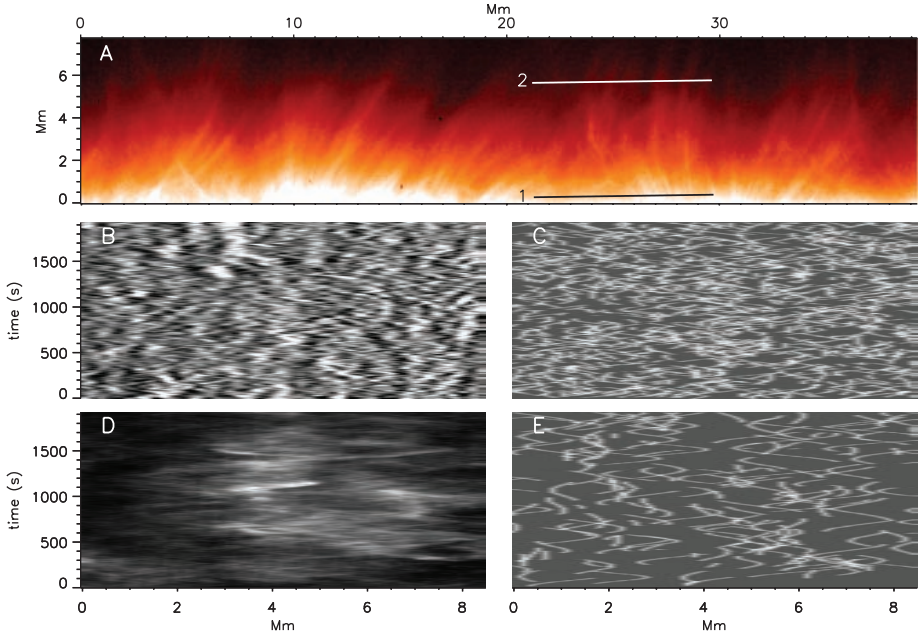


Fig. 2. Illustration of the ubiquity of Alfvén waves in the chromosphere. (**A**) A Hinode SOT Ca II H 3968 Å image showing how thin spicules that outline the magnetic field dominate the chromosphere. (**B**) A space-time plot [along the cut labeled 1 in (A)] of the Ca intensity processed to enhance 200-km-wide structures. The plot is dominated by a multitude of criss-crossed short linear tracks caused by spicular motion transverse to the magnetic field direction. (**D**) A similar cut for the line labeled 2 in (A). Image enhancement is unnecessary at these heights because of the smaller number of spicules. Similar linear tracks are visible, as well as swings. The general characteristics (linear tracks and swings) of (B) and (D) are well reproduced by cuts that are generated from Monte Carlo simulations [(**C**) and (**E**)] (*12*) in which spicules carry Alfvén waves.

[1]Lockheed Martin Solar and Astrophysics Laboratory, 3251 Hanover Street, Organization ADBS, Building 252, Palo Alto, CA 94304, USA. [2]Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Suite 400, Boulder, CO 80302, USA. [3]High Altitude Observatory, National Center for Atmospheric Research, Post Office Box 3000, Boulder, CO 80307, USA. [4]Institute of Theoretical Astrophysics, University of Oslo, Post Office Box 1029, Blindern, 0315 Oslo, Norway. [5]National Astronomical Observatory, Mitaka, Tokyo 181-8588, Japan. [6]Institute of Space and Astronautical Science/Japan Aerospace Exploration Agency, Sagamihara, Kanagawa 229-8510, Japan. [7]Kwasan and Hida Observatories, Kyoto University, Yamashina, Kyoto 607-8471, Japan.

*To whom correspondence should be addressed. E-mail: bdp@lmsal.com

illustrated in Fig. 2D. This is well reproduced by our Monte Carlo simulation (Fig. 2E) with fewer and longer-lived spicules (with lifetimes distributed around 100 s instead of around 40 s).

We also used Monte Carlo simulations to determine the properties of the observed Alfvén waves by comparing the maximum velocity and maximum transverse displacement of space-time tracks perpendicular to observed spicules with similar properties from simulated spicules (*12*). Our observations show that most spicules undergo some transverse displacement, usually between 200 and 500 km, and maximum transverse velocities of about 10 to 30 km/s, with a peak in the distribution around 15 km/s (Fig. 3). These observed distributions of maximum velocity and displacement are best reproduced by Monte Carlo simulations in which spicules carry Alfvén waves with velocity amplitudes that have a Gaussian distribution centered at 20 ± 5 km/s and a uniform distribution of periods between 150 and 350 s. The best fit agrees in almost all bins of displacement and velocity to within the estimated error bars (*12*). The fact that the observed distributions do not show an excess at 0 km displacement and 0 km/s velocity implies that our data are compatible with a model in which essentially all of the observed chromospheric features carry Alfvén waves. In addition, these comparisons are quite sensitive to the input velocity amplitude distribution: Higher average velocities (such as 25 km/s) or lower average velocities (such as 15 km/s) do not match the observed velocity amplitude distribution at either end (fig. S3).

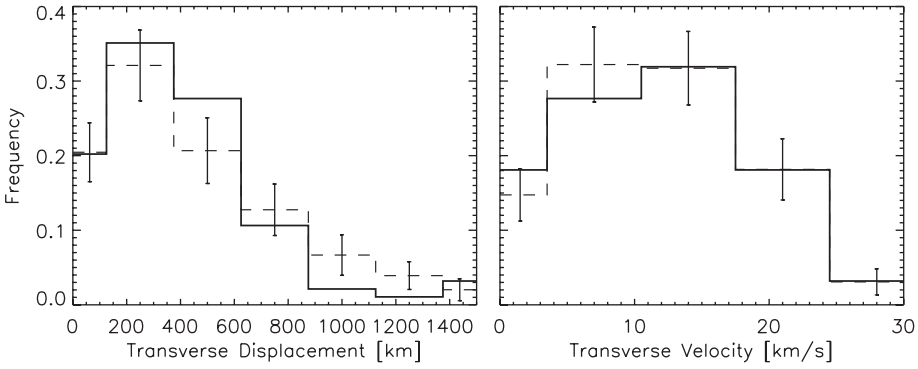


Fig. 3. Comparison between observed and simulated transverse displacements and velocity amplitudes. The left panel shows the distribution of measured transverse displacements of 94 observed spicules [full line (*12*)] and from a Monte Carlo simulation (dashed line) in which spicules carry Alfvén waves with periods randomly chosen from a uniform distribution between 150 and 350 s and velocity amplitudes from a Gaussian distribution around 20 ± 5 km/s. The simulated and observed distributions agree well, especially when taking into account the errors introduced by the poor statistics because of the low number of spicules measured [dashed error bars (*12*)]. The right panel shows a similar comparison for the observed and simulated transverse velocity amplitudes. The agreement between observed and simulated distributions indicates that our data are compatible with ubiquitous Alfvén waves that affect most observed spicules.

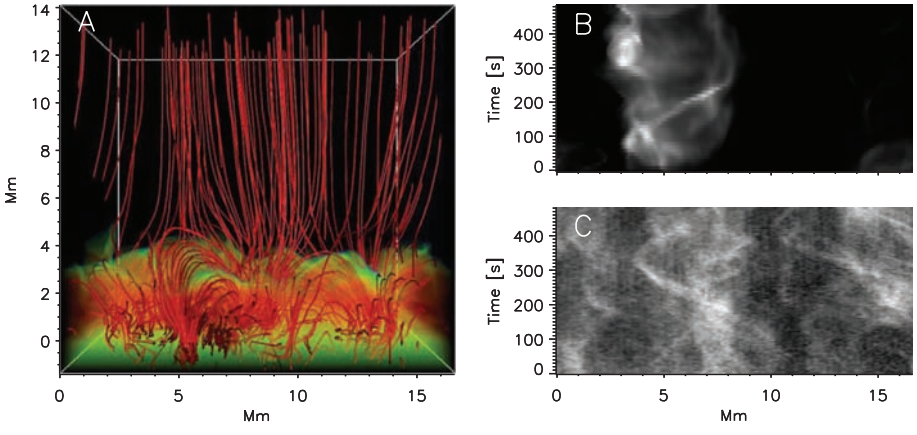


Fig. 4. Comparison between observations and simulations of Alfvén waves. (**A**) A snapshot from a self-consistent 3D radiative MHD simulation ranging from the convection zone up to the corona. Analysis shows that the field lines (red lines) in the corona, transition region, and chromosphere are continuously shaken and carry Alfvén waves (movie S5). The coloring shows the plasma temperature from lower chromospheric values (red) to higher transition-region temperatures (green). A space-time cut of the Ca II H 3968 Å synthetic intensity [(**B**), from simulations] shows similar half-sinusoidal and linear tracks as a space-time plot from the observations (**C**).

The wave periods are more difficult to determine because the lifetimes of the spicules delineating the waves are generally much shorter than the wave period. However, visual comparisons between the observations and simulated data limit the periods to between 100 and 500 s: Waves with very short periods on the order of 100 s or very long periods on the order of 500 s do not fit the data very well (fig. S4) (*13*). Our observations suggest that very long-lived macrospicules (lifetimes of >10 min and heights >10,000 km) show some evidence of Alfvén waves with longer periods between 300 and 600 s (fig. S5).

To study the propagation and impact on the atmosphere of these Alfvén waves, we turned to advanced three-dimensional (3D) radiative MHD simulations of a region on the Sun encompassing its convection zone, photosphere, chromosphere, transition region, and corona (*12*, *14*, *15*). Our simulations showed ubiquitous Alfvén waves with properties similar to those of the waves found in our observations, as illustrated in the xt cut of synthetic Ca II H emission at a height of 4.8 megameters (Mm = 10^6 m) in the MHD simulations (Fig. 4). Detailed analysis of the simulations makes it unequivocally clear that these are volume-filling Alfvén waves, with the magnetic field lines swaying back and forth as the waves pass (movie S5 and fig. S6). We do not see evidence for stable waveguides or MHD kink-mode waves. Our simulations show that despite reflections in the chromosphere and transition region, Alfvén waves with significant amplitudes propagate into the corona from below, with transmission coefficients from the chromosphere into the corona on the order of 3 to 15%. Similar values have also been obtained by 1.5D modeling (*5*–*7*, *16*–*19*).

This means that these waves are very important for the energy balance of the solar outer atmosphere. We estimate the energy flux in the chromosphere $E = \rho v^2 v_A = 4$ to 7 kW m^{-2}, with the Alfvén speed $v_A = B/\sqrt{(\mu_0 \rho)}$, with magnetic permeability μ_0. This estimate is based on the observed velocity amplitude $v \approx 20$ km/s, conservative values for the spicule mass density ($\rho = 2.2 \times 10^{-11}$ to 4×10^{-10} kg m^{-3}) (*20*), and measured spicule magnetic fields $B \approx 10^{-3}$ T (*21*). The implied Alfvén speeds are on the order of 45 to 200 km/s, which is compatible with our estimates of larger than 50 to 200 km/s based on the observed periods and minimum wavelength of the waves (*12*). The energy flux that reaches the corona is thus on the order of 120 W m^{-2} for a transmission coefficient of 3%, which is on the low end of what is expected theoretically. This energy flux is large enough to supply the energy necessary to heat the quiet Sun corona and/or drive the solar wind (~100 W m^{-2}) (*22*). This value is fully compatible with recent measurements with the coronal multichannel polarimeter (CoMP) instrument of coronal Alfvén waves with periods that are similar to those we report here (*23*). Although CoMP directly measures amplitudes

that are only on the order of 0.5 km/s with a corresponding energy flux of ~0.01 W m^{-2}, Tomczyk *et al*. (*23*) point out that line-of-sight superposition in the optically thin corona can reduce the directly observed wave amplitudes and hide the remaining wave energy flux in the large nonthermal linewidths observed with CoMP.

Once in the corona, Alfvén waves propagate quite freely, and although modelers disagree on which dissipation mechanism dominates, they agree that Alfvén waves with an energy flux of 100 W m^{-2} are vigorous enough to launch the solar wind when, inevitably, their energy is thermalized and their momentum flux is added to the wind (*5*, *6*). Our observations of vigorous low-frequency waves obviate the need for high-frequency models in which Alfvén waves are assumed to be generated by reconnection processes in the low corona and dissipated through resonant absorption as the wave frequency becomes equal to the gyrofrequencies of the various ions in the plasma (*2*, *3*, *24*, *25*). The observational support for such models has recently been challenged (*26*, *27*). Our data, on the contrary, strongly support the recent models that are based on the dissipation of low-frequency waves; for example, because of self-interference from reflection (*4*, *28*), compressible effects (*7*), or parametric decay (*29*, *30*). Both the observed amplitudes and periods are consistent with those in comprehensive simulations that describe the generation, propagation, and dissipation of Alfvén waves from the photosphere to Earth orbit (*5*, *6*).

References and Notes

1. J. W. Belcher, S. Olbert, *Astrophys. J.* **200**, 369 (1975).
2. W. I. Axford *et al*., *Space Sci. Rev.* **87**, 25 (1999).
3. C.-Y. Tu, E. Marsch, *J. Geophys. Res.* **106**, 8233 (2001).
4. W. H. Matthaeus, G. P. Zank, S. Oughton, D. J. Mullan, P. Dmitruk, *Astrophys. J.* **523**, L93 (1999).
5. A. Verdini, M. Velli, *Astrophys. J.* **662**, 669 (2007).
6. S. R. Cranmer, A. A. van Ballegooijen, R. J. Edgar, *Astrophys. J.* **171** (suppl.), 520 (2007).
7. T. K. Suzuki, S.-i. Inutsuka, *J. Geophys. Res.* **111**, A06101 (2006).
8. S. Tsuneta *et al*., *Sol. Phys.*, http://arxiv.org/abs/0711.1715.
9. T. Kosugi, *et al., Sol. Phys.* **243**, 3 (2007).
10. B. De Pontieu *et al*., *Publ. Astron. Soc. Jpn.* **59**, S655 (2007).
11. For a discussion on MHD kink-mode waves, see the supporting online material (SOM).
12. Methods are available as supporting material on *Science* Online.
13. For a discussion on what may determine the periods, see the SOM.
14. V. H. Hansteen, B. Gudiksen, *ESA Spec. Pub.* **592**, 87 (2005).
15. V. H. Hansteen, M. Carlsson, B. Gudiksen, *Astron. Soc. Pac. Conf. Ser.* **368**, 107 (2007).
16. T. Kudoh, K. Shibata, *Astrophys. J.* **514**, 493 (1999).
17. J. V. Hollweg, *Sol. Phys.* **56**, 305 (1978).
18. J. V. Hollweg, *Sol. Phys.* **70**, 25 (1981).
19. J. V. Hollweg, S. Jackson, D. Galloway, *Sol. Phys.* **75**, 35 (1982).
20. J. M. Beckers, *Sol. Phys.* **3**, 367 (1968).
21. J. Trujillo Bueno, L. Merenda, R. Centeno, M. Collados, E. Landi Degl'Innocenti, *Astrophys. J.* **619**, L191 (2005).
22. V. H. Hansteen, E. Leer, *J. Geophys. Res.* **100**, 21577 (1995).
23. S. Tomczyk *et al*., *Science* **317**, 1192 (2007).
24. S. R. Cranmer, G. B. Field, J. L. Kohl, *Astrophys. J.* **518**, 937 (1999).
25. S. R. Cranmer *et al*., *Astrophys. J.* **511**, 481 (1999).
26. N.-E. Raouafi, S. K. Solanki, *Astron. Astrophys.* **412**, 271 (2003).
27. N.-E. Raouafi, J. W. Harvey, S. K. Solanki, *Astrophys. J.* **658**, 643 (2007).
28. M. Velli, *Astron. Astrophys.* **270**, 304 (1993).
29. F. Pruneti, M. Velli, *ESA Spec. Pub.* **404**, 623 (1997).
30. L. Del Zanna, M. Velli, P. Londrillo, *Astron. Astrophys.* **367**, 705 (2001).

31. This research was supported by NASA grants NNG06GG79G, NNG06GC89G, and NM07AA01C (Hinode/SOT); NSF grant ATM-0541567; and the Research Council of Norway through grant 170935/V30 and through grants of computing time from the Programme for Supercomputing. B.D.P. thanks the Institute of Theoretical Astrophysics/Oslo group and S.W.M. for excellent hospitality. We thank the Hinode/SOT team for building an excellent instrument. Hinode is a Japanese mission developed and launched by the Institute of Space and Astronautical Science/Japan Aerospace Exploration Agency (ISAS/JAXA), with the National Astronomical Observatory of Japan (NAOJ) as domestic partner and NASA and the Science and Technology Facilities Council (UK) as international partners. It is operated by these agencies in cooperation with the European Space Agency and Norwegian Space Centre (Norway). The SOT was developed jointly by NAOJ, the Lockheed Martin Solar and Astrophysics Laboratory, ISAS/JAXA, NASA, the High Altitude Observatory, and Mitsubishi Electric Corporation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1574/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
References
Movies S1 to S5

22 May 2007; accepted 13 November 2007
10.1126/science.1151747

REPORT

Coronal Transverse Magnetohydrodynamic Waves in a Solar Prominence

T. J. Okamoto,[1,2]* S. Tsuneta,[1] T. E. Berger,[3] K. Ichimoto,[1] Y. Katsukawa,[1] B. W. Lites,[4] S. Nagata,[2] K. Shibata,[2] T. Shimizu,[5] R. A. Shine,[3] Y. Suematsu,[1] T. D. Tarbell,[3] A. M. Title[3]

Solar prominences are cool 10^4 kelvin plasma clouds supported in the surrounding 10^6 kelvin coronal plasma by as-yet-undetermined mechanisms. Observations from Hinode show fine-scale threadlike structures oscillating in the plane of the sky with periods of several minutes. We suggest that these represent Alfvén waves propagating on coronal magnetic field lines and that these may play a role in heating the corona.

Solar prominences are classified as either quiescent or active region (AR), the latter referring to material suspended above sunspot magnetic regions. Quiescent prominences often exist for many weeks at high solar latitudes, whereas AR prominences can be dynamic and short-lived. They are the most enigmatic of solar structures supported by coronal magnetic field lines, sometimes erupting as the source of coronal mass ejections, large-scale eruptions of plasma from flaring solar active regions, that can have major impacts on the terrestrial magnetic environment. Recent ground-based observations have revealed that AR prominences have numerous small threadlike features (*1*), with continuous flow of material along the threads (*2*–*9*). Observations from space (*10*, *11*) confirm these findings and show additional dynamics related to coronal structure.

We report Hinode Solar Optical Telescope (SOT) (*12*, *13*) observations of an AR prominence in a 0.3-nm broadband region centered at 396.8 nm, the H-line spectral feature of singly ionized calcium (Ca II). Radiation in this bandpass typically has a temperature of less than 20,000 K.

We obtained over 1 hour of continuous SOT images of NOAA AR 10921 on the west solar limb on 9 November 2006. The images show a multithreaded AR prominence suspended above the main sunspot (Fig. 1). Although no simultaneous Hα images were taken, the Ca II H-line prominence structures are consistent with the structures seen in lower-resolution Hα observations (*14*). The Ca II H-line movie (movie S1) shows ubiquitous continuous horizontal motions along the prominence threads. The origins of these flows remain unknown. Some of the flows had constant speeds of about 40 km s^{-1}, whereas others accelerated monotonically in a more complicated fashion.

The Hinode SOT movies also reveal that many of the threads in the prominence underwent vertical (i.e., in the plane of the sky) oscillatory motions (Fig. 2) at periods of 130 to 250 s. The

[1]National Astronomical Observatory (NAOJ), Mitaka, Tokyo 181-8588, Japan. [2]Kwasan and Hida Observatories, Kyoto University, Yamashina, Kyoto 607-8471, Japan. [3]Lockheed Martin Solar and Astrophysics Laboratory, B/252, 3251 Hanover Street, Palo Alto, CA 94304, USA. [4]High Altitude Observatory, National Center for Atmospheric Research, Post Office Box 3000, Boulder CO 80307–3000, USA. [5]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency (ISAS/JAXA), Sagamihara, Kanagawa 229-8510, Japan.

*To whom correspondence should be addressed. E-mail: joten.okamoto@nao.ac.jp

vertical oscillatory motions were coherent over lengths as long as 16,000 km. One thread (Fig. 3A) had a vertical width of 660 km and an oscillation period of 240 s. Comparison of the oscillation phase at various horizontal positions (Fig. 3, B to F) reveals that the thread oscillated in phase along its entire length. The vertical oscillation amplitudes of all threads we saw ranged from 400 to 1800 km, and thread widths were between 430 and 660 km (Table 1). Because we cannot determine the angle between the plane of oscillation and the line of sight to the prominence threads, the oscillation amplitudes as well as the horizontal velocities are minimum estimates.

The observed vertical oscillatory motion is most likely due to propagating or standing Alfvén waves on the horizontal magnetic field lines that compose the prominence. An alternate hypothesis is longitudinal plasma motion along helical field lines. Helical field line structure has been observed in many previous prominence studies (*15*–*17*); however, the helical field line hypothesis cannot explain the several cases we found of synchronous vertical oscillation of entire threads, for example, Fig. 3. Alternately, if the observed threads are thin magnetic flux tubes, the observed oscillations may be fast magneto-acoustic kink modes propagating along the tubes (*18*, *19*). The observations shown here lack line-of-sight Doppler velocity measurements, so we cannot determine the exact oscillation mode at this point.

These field lines connect to the source regions in the photosphere (*7*), where they are excited by the wide-spectrum p-mode oscillations originating in the convection zone. Each field line oscillated independently, as seen in the Hinode movies. The synchronous oscillation along entire threads shown in Fig. 3 implies that we cannot distinguish the time difference at the minimum and maximum amplitudes all along the thread. The uncertainty in phase of the oscillation is no more than 1/16 of the oscillation period, so we estimate that the minimum wavelength of the oscillation is 16 × 16,000 ≈ 250,000 km. The wave speed is estimated to be >1050 km s^{-1} for the average oscillation period of ~240 s. If we assume that the plasma density is 10^{10} cm^{-3} (*20*), the implied magnetic field strength is ~50 G for the propagating Alfvén wave, in agreement with measurements and models of active region prominence magnetic fields (*21*, *22*). The Poynting flux carried by the observed waves is then estimated

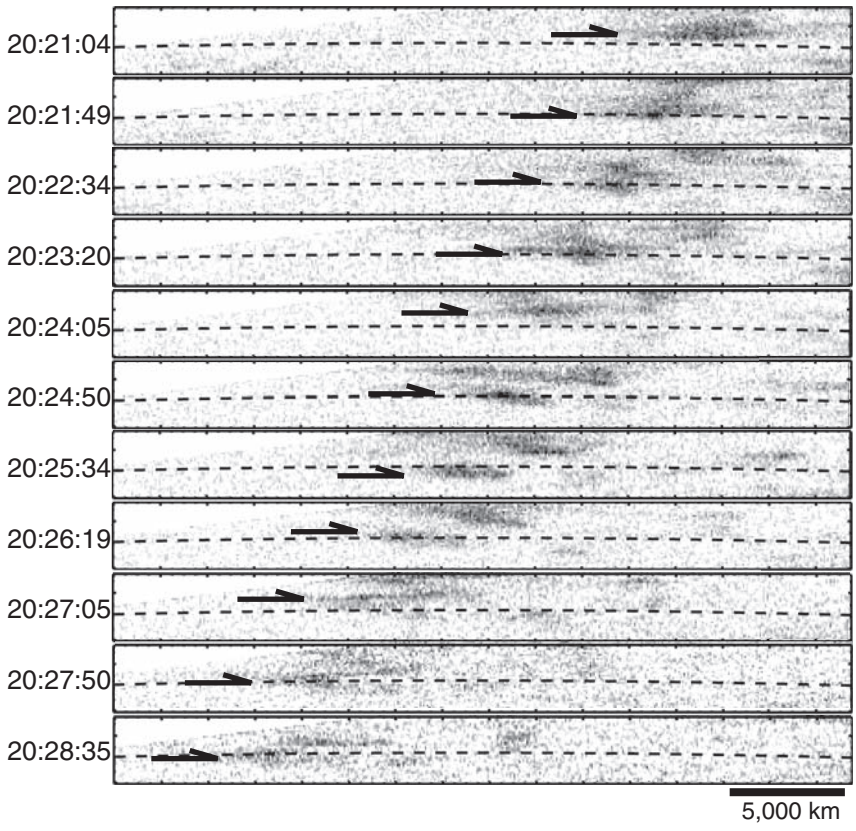


Fig. 2. Example of a vertical oscillation of a single prominence thread. A small field of view is extracted from the larger field shown in Fig. 1 and shown in negative contrast. Tick marks have a spacing of 2000 km on the Sun, and UT time for each image is denoted on the left. The dashed line in each image indicates an approximately constant height above the photosphere. This oscillating thread has a length of about 3600 km and width of 430 km. Steady flow at 39 km s^{-1} along the thread is evident. The vertical amplitude of the oscillation is about 900 km with a period of 174 s. The vertical speed is about 10 km s^{-1}.

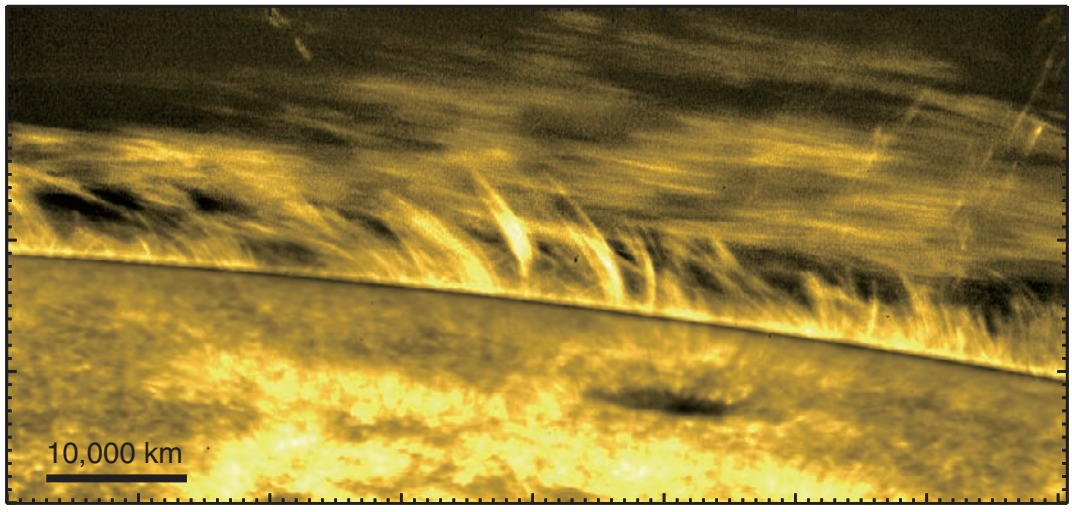


Fig. 1. High-resolution image on the solar limb obtained with SOT aboard Hinode. This observation was performed with a cadence of 15 s from 19:33 to 20:44 UT on 9 November 2006. Tick marks have a spacing of 1000 km on the Sun. A radial density filter is applied to show the brighter photosphere and the fainter coronal structures in the same image. The main sunspot of NOAA AR 10921 as well as the trailing bright plage areas are visible on the disk. Above the limb, ubiquitous vertical spicules are seen below the horizontal threads of the AR prominence. The cloudlike prominence structure is located 10,000 to 20,000 km above the visible limb and exhibits a very complex fine structure with predominant horizontal threadlike features. The intensity of the prominence in Ca II H-line radiation is about 1% of the on-disk photosphere.

to be $\rho v^2 V_A \sim 2.0 \times 10^6$ erg s^{-1} cm^{-2} (where ρ is the density, v is the velocity amplitude, and V_A is the Alfvén speed), a lower limit based on the minimum estimates of observed tangential velocities. Given a suitable dissipation mechanism, this flux is sufficient to heat coronal loops with lengths longer than the estimated oscillation wavelength (*23*).

In transiting from the photosphere to the chromosphere, the large density decrease with height results in rapid increases in the acoustic and Alfvén wave propagation speeds. This effective discontinuity causes waves with periods longer than a cutoff period determined by the thermodynamic conditions and magnetic field strength in the atmosphere to be reflected before reaching the coronal heights of active region prominences. Because we find a typical oscillation period ~ 240 to 250 s, we can infer that the Alfvén cutoff period is longer than about 4 min for this prominence structure.

The limited field of view of our data prevents us from determining the lengths of the field lines threading the prominence. However, the frequency spectrum of the observed waves can be used to investigate the length of the field lines. Open field lines have a distinct wide spectrum above the Alfvén cutoff frequency, whereas shorter closed loops show multiple discrete resonances as a function of loop length. The estimated Alfvén speed and the mean observed period implies a minimum length of 250,000 km if this is the standing wave in a closed-loop system.

Previous observations of waves in the solar corona include Doppler velocity and periodic intensity oscillations in coronal loops as well as flare-generated transversal displacements of active region loops (*24*–*28*). Those oscillations are examples of magneto-acoustic waves propagating out from photospheric source sites. Alfvén waves in coronal loops and prominences are also claimed to account for spectroscopic observations of nonthermal line widths in coronal emission lines (*29*, *30*).

Table 1. Properties of moving threads with vertical oscillations. Asterisks mean averaged velocity.

	Length (km)	Width (km)	Horizontal velocity (km^{-1})	Vertical oscillation period (s)	Vertical oscillation width (km)	Height from the limb (km)
1	3600	430	39	174 ± 25	904	18,300
2	16,000	660	15	240 ± 30	1113	12,400
3	6700	580	39	230 ± 87	909	14,700
4	2200	360	46*	180 ± 137	435	19,000
5	3500	430	45*	135 ± 21	408	14,300
6	1700	510	25*	250 ± 17	1771	17,200

References and Notes

1. Y. Lin, O. Engvold, L. Rouppe van der Voort, J. E. Wiik, T. E. Berger, *Sol. Phys.* **226**, 239 (2005).
2. J. C. Vial, P. Gouttebroze, G. Artzner, P. Lemaire, *Sol. Phys.* **61**, 39 (1979).

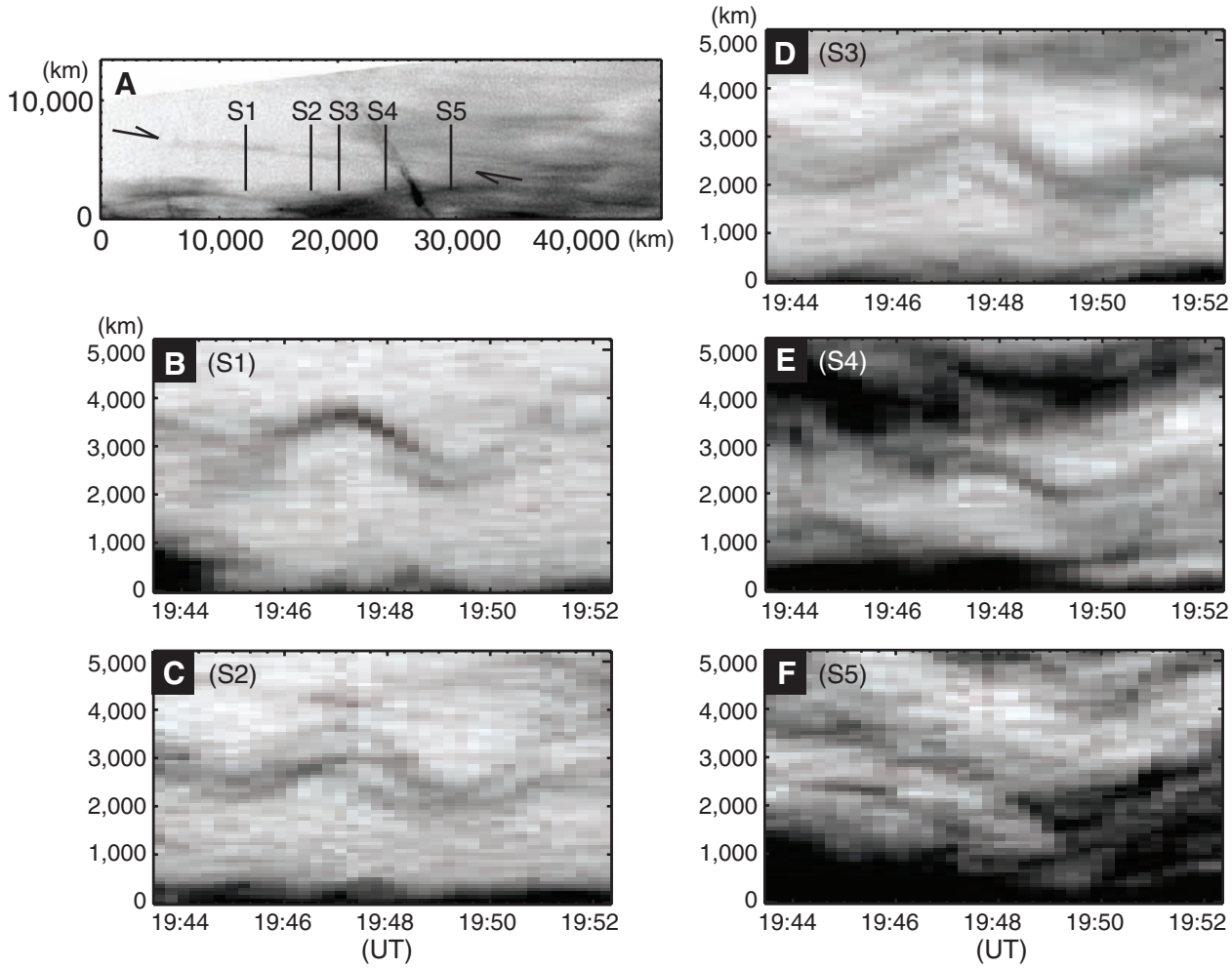


Fig. 3. Example of a prominence thread undergoing synchronous oscillation along its entire length. (**A**) The long thread extending ~16,000 km. Lines S1 to S5 indicate the locations of height versus time plots shown in (B to F). The arrows indicate both sides of the long thread oscillating synchronously. (**B** to **F**) Height-time plots (shown in negative contrast) for the locations indicated in (A). Maximum and minimum amplitudes occur at nearly the same time for all locations.

3. J. M. Malherbe, B. Schmieder, E. Ribes, P. Mein, *Astron. Astrophys.* **119**, 197 (1983).
4. G. Simon, B. Schmieder, P. Démoulin, A. I. Poland, *Astron. Astrophys.* **166**, 319 (1986).
5. B. Schmieder, M. A. Raadu, J. E. Wiik, *Astron. Astrophys.* **252**, 353 (1991).
6. S. F. Martin, *Sol. Phys.* **182**, 107 (1998).
7. J. B. Zirker, O. Engvold, S. F. Martin, *Nature* **396**, 440 (1998).
8. J. Chae, C. Denker, T. J. Spirock, H. Wang, P. R. Goode, *Sol. Phys.* **195**, 333 (2000).
9. J. T. Karpen, S. K. Antiochos, J. A. Klimchuk, *Astrophys. J.* **637**, 531 (2006).
10. C. J. Schrijver *et al.*, *Sol. Phys.* **187**, 261 (1999).
11. S. Patsourakos, J. C. Vial, *Sol. Phys.* **208**, 253 (2002).
12. T. Kosugi *et al.*, *Sol. Phys.* **243**, 3 (2007).
13. S. Tsuneta *et al.*, http://arxiv.org/abs/0711.1715.
14. Y. Suematsu, R. Yoshinaga, N. Terao, T. Tsubaki, *Publ. Astron. Soc. Jpn.* **42**, 187 (1990).
15. M. Kuperus, M. A. Raadu, *Astron. Astrophys.* **31**, 189 (1974).
16. J. L. Leroy, V. Bommier, S. Sahal-Bréchot, *Astron. Astrophys.* **131**, 33 (1984).
17. T. Hirayama, *Sol. Phys.* **100**, 415 (1985).
18. V. M. Nakariakov, E. Verwichte, *Living Rev. Sol. Phys.* **2**, 3 (2005).
19. A. J. Díaz, R. Oliver, J. L. Ballester, *Astrophys. J.* **580**, 550 (2002).
20. T. Hirayama, in *Coronal and Prominence Plasma*, A. I. Poland, Ed., NASA Conference Publication No. 2442 (NASA, Washington, DC, 1986).
21. E. Tandberg-Hanssen, J. M. Malville, *Sol. Phys.* **39**, 107 (1974).
22. E. Wiehr, G. Stellmacher, *Astron. Astrophys.* **247**, 379 (1991).
23. G. L. Withbroe, R. W. Noyes, *Annu. Rev. Astron. Astrophys.* **15**, 363 (1977).
24. I. De Moortel, J. Ireland, A. W. Hood, R. W. Walsh, *Astron. Astrophys.* **387**, L13 (2002).
25. C. J. Schrijver, M. J. Aschwanden, A. M. Title, *Sol. Phys.* **206**, 69 (2002).
26. M. J. Aschwanden, B. De Pontieu, C. J. Schrijver, A. M. Title, *Sol. Phys.* **206**, 99 (2002).
27. D. Banerjee, R. Erdélyi, R. Oliver, E. O'Shea, *Sol. Phys.* **tmp**, 136 (2007).
28. Y. Lin, O. Engvold, L. H. M. Rouppe van der Voort, M. van Noort, *Sol. Phys.* **tmp**, 71 (2007).
29. S. Koutchmy, Y. D. Zugzda, V. Locans, *Astron. Astrophys.* **120**, 185 (1983).
30. H. Hara, K. Ichimoto, *Astrophys. J.* **513**, 969 (1999).

31. The authors thank H. Shibahashi, T. Sekii, R. Erdélyi, and V. Nakariakov for comments. Hinode is a Japanese mission developed and launched by ISAS/JAXA, with NAOJ as domestic partner and NASA and Science and Technology Facilities Council (STFC) (UK) as international partners. It is operated by these agencies in cooperation with European Space Agency and Norwegian Space Centre (Norway). This work was carried out at the NAOJ Hinode science center, which was supported by the Grant-in-Aid for Creative Scientific Research, the Basic Study of Space Weather Prediction (head investigator, K.S.) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, donation from Sun Microsystems Incorporated, and NAOJ internal funding. The National Center for Atmospheric Research is sponsored by NSF. T.J.O. is supported by research fellowships from the Japan Society for the Promotion of Science for Young Scientists.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1577/DC1
Movie S1

21 May 2007; accepted 6 November 2007
10.1126/science.1145447

REPORT

Evidence for Alfvén Waves in Solar X-ray Jets

J. W. Cirtain,[1,2]* L. Golub,[1] L. Lundquist,[1] A. van Ballegooijen,[1] A. Savcheva,[1] M. Shimojo,[3] E. DeLuca,[1] S. Tsuneta,[4] T. Sakao,[5] K. Reeves,[1] M. Weber,[1] R. Kano,[4] N. Narukage,[5] K. Shibasaki[3]

Coronal magnetic fields are dynamic, and field lines may misalign, reassemble, and release energy by means of magnetic reconnection. Giant releases may generate solar flares and coronal mass ejections and, on a smaller scale, produce x-ray jets. Hinode observations of polar coronal holes reveal that x-ray jets have two distinct velocities: one near the Alfvén speed (~800 kilometers per second) and another near the sound speed (200 kilometers per second). Many more jets were seen than have been reported previously; we detected an average of 10 events per hour up to these speeds, whereas previous observations documented only a handful per day with lower average speeds of 200 kilometers per second. The x-ray jets are about 2×10^3 to 2×10^4 kilometers wide and 1×10^5 kilometers long and last from 100 to 2500 seconds. The large number of events, coupled with the high velocities of the apparent outflows, indicates that the jets may contribute to the high-speed solar wind.

The solar corona provides an opportunity to study the interactions of high-temperature electrically conducting gas, plasma, and a dynamic magnetic field. The constant emergence and cancellation of the magnetic field create a multitude of energetic changes in magnetic topology that can inject enormous amounts of energy into the plasma. It is thought that magnetic reconnection is involved in releasing energy to produce solar flares and coronal mass ejections (CMEs). X-ray jets, in which a burst of hot plasma is driven into the solar corona along an open magnetic field line, are thought to be a different manifestation of the reconnection process. It appears as though at least some fraction of these outflows has sufficient kinetic energy to leave the corona and propagate into the inner heliosphere.

Two types of outflows are possible during the post–magnetic reconnection phase of a jet. In the first case, an outflow at the local sound speed (v_c), resulting from energy deposition that rapidly heats the dense chromospheric plasma, expands into the overlying low-pressure corona. The sound speed is governed by the equation

$$v_c = \left(\frac{2\gamma kT^{1/2}}{m_p}\right)^{1/2} \quad (1)$$

Here, T is the temperature at the location of energy deposition, k is the Boltzmann constant, γ is the ratio of specific heat capacities, and m_p is the proton mass. The average temperature of jets and the related footpoint flares is 6 million K (*1*, *2*), a temperature estimate consistent with our observations, which also show that the loop density in the coronal hole before the jet is 3×10^8 cm^{-3}. The initial velocity of the conduction front would be 400 km s^{-1}.

The second case is when plasma is accelerated by the formation of an Alfvén wave during the relaxation of the magnetic field, and the plasma is forced to flow out along the field at about the Alfvén speed, v_a, given by

$$v_a = B_0/\sqrt{4\pi\rho} \quad (2)$$

where B_0 is the magnetic field strength, and ρ is the mass density. Assuming that the magnetic field is of the order of 10 gauss, the Alfvén velocity of the plasma before the subsequent evaporation should be ~ 1000 km s^{-1}.

These observations of coronal jets provide useful insight into the formation of hot, collimated, high-velocity outflows, which are likely a large-scale contributor to the mass loading of the fast solar wind. The first models for the existence of the solar wind were provided by Parker and Chapman (*3*, *4*). Previous work (*5*–*9*) has suggested that Alfvén waves may play an important role in driving the solar wind. Using observations from the NASA mission Ulysses, Wang (*10*) found that the fast solar wind originated from the polar coronal holes and was nearly continuously present, but there was no direct evidence for the mechanism producing the fast wind.

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA. [2]Marshall Space Flight Center, National Aeronautics and Space Administration (NASA) VP62, Huntsville, AL 35812, USA. [3]Nobeyama Solar Radio Observatory, Nobeyama, Nagano 384–1305, Japan. [4]National Astronomical Observatory of Japan (NAOJ), Mitaka, Tokyo 181–8588, Japan. [5]Institute of Space and Astronautical Science, (ISAS), Japan Aerospace Exploration Agency (JAXA), Sagamihara, Kanagawa 229–8510, Japan.

*To whom correspondence should be addressed. E-mail: Jonathan.W.Cirtain@nasa.gov

X-ray jets have been identified in polar coronal holes, in "quiet" Sun, and from within active regions (*1*, *2*, *10*–*14*). Shibata *et al.* (*2*) found that the jets were a transient x-ray source with essentially collimated motion outward along the coronal magnetic field from the initiation site. The observed outflow velocity in these studies was typically 200 to 600 km s^{-1}. These jets were reported to have lengths of 1×10^5 to 10×10^5 km and collimated widths of 1×10^4 km. Other observations of polar jets by means of instruments such as the High-Resolution Telescope and Spectrograph (HRTS) (*13*, *14*) have reported that extreme ultraviolet jets, so-called chromospheric jets, do not have physical characteristics similar to the x-ray jets reported by Shibata and others or to the x-ray jets studied in this report. The relationship between the lower-temperature jets (observed with emission lines formed near 1.0×10^4 K) and x-ray jets (formed above 2.0×10^6 K) remains unclear.

Hinode (formerly Solar-B) is in polar orbit about Earth, following the day-night terminator and thus providing nearly continuous observations of the Sun. Here we report Hinode X-ray Telescope (XRT) observations of polar coronal hole jets and show that at least some jets have two velocity components. According to current theories of magnetic reconnection, an Alfvén wave should be generated by the reconnected magnetic field line as it proceeds from a highly curved geometry to a relaxed configuration. This Alfvén wave could drive plasma along the field at speeds of 600 to 1000 km s^{-1}, depending on the local plasma density and field strength. We have observed several such outflow (radial) velocities for some large jets. The energy released by reconnection will subsequently heat the plasma, which expands into the corona at v_c. We have also observed this component of the process and can clearly differentiate it from the high-speed component.

XRT collected >9000 images during 10 different 6- to 8-hour continuous observations [Figs. 1 and 2, supporting online material (SOM) text, and SOM movies S1 and S2]. Both north and south polar coronal holes were studied. XRT images had a 1024-by-512 arc second field-of-view (7.68×10^5 by 3.84×10^5 km), with one image taken every 30 s. The observed jets are typically 2×10^3 to 2×10^4 km wide and greater than 1×10^5 km long.

We define the axis of a jet to be a line beginning at the initiation site of the jet and extending in the outflow direction. We determined the x-ray intensity along this line from many sequential images. This measure of intensity along the jet axis is plotted as a column and shows the x-ray intensity variation along the jet axis at one instant in time. By "stacking" these columns from sequential images along the *x* axis of the plot, a representation of the intensity variation in distance and time is created (Fig. 3A). The slope of an intensity front in this type of plot determines the velocity of the outflowing plasma.

Using this technique, we examined four jets in detail. There are multiple velocity components for each of the jets. One component of the jet velocity is consistent with the previously reported spatio-temporal average velocity of ≈ 200 km s^{-1} (*1*). However, a much higher velocity is also observed, roughly ~ 800 km s^{-1} at the start of each event. We interpret this as evidence for material being ejected at v_a during the relaxation of the magnetic field

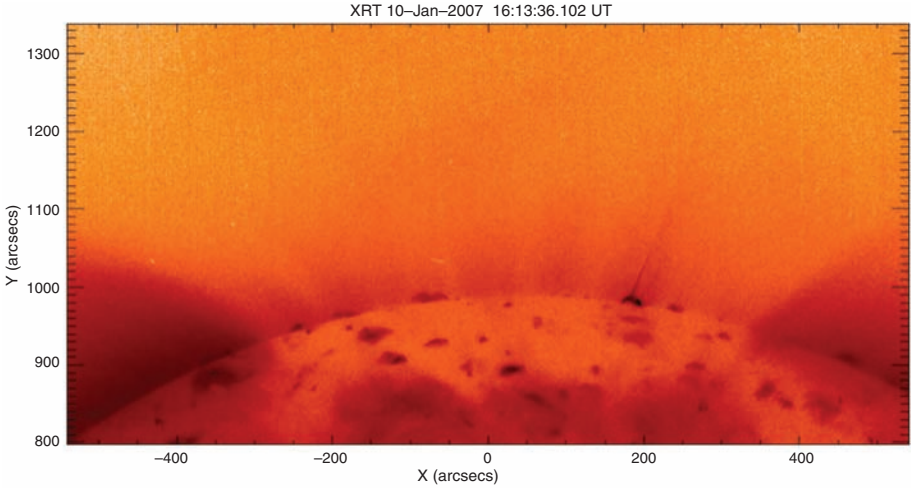


Fig. 1. Hinode XRT false-color image of the north polar coronal hole. A typical jet is seen in the center of this image (movie S2).

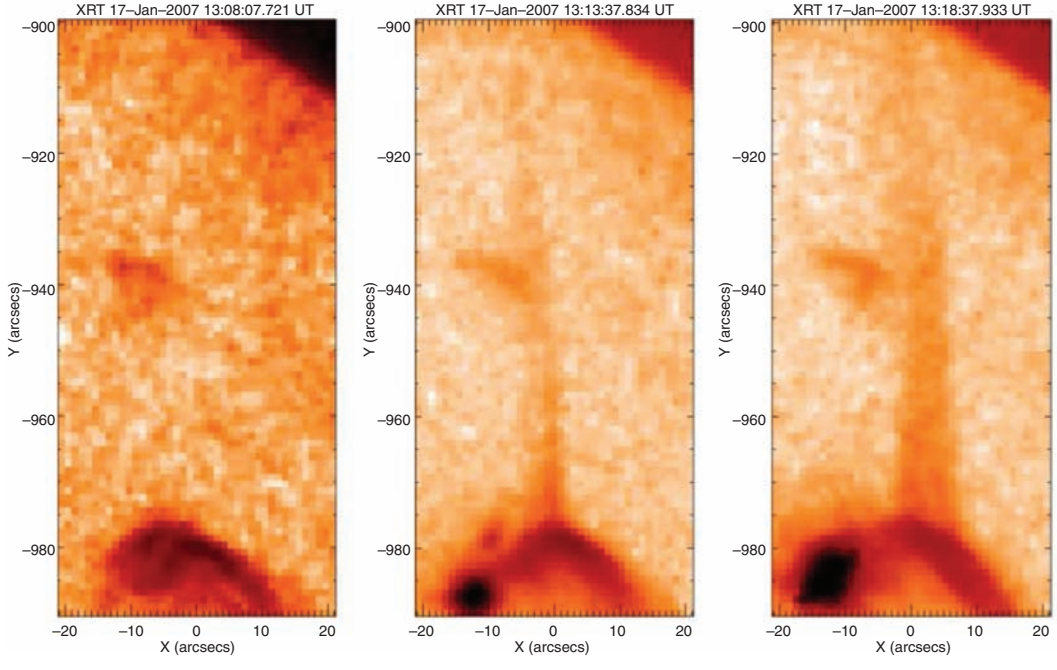


Fig. 2. Hinode XRT false-color images of three stages of a jet's evolution (movie S1).

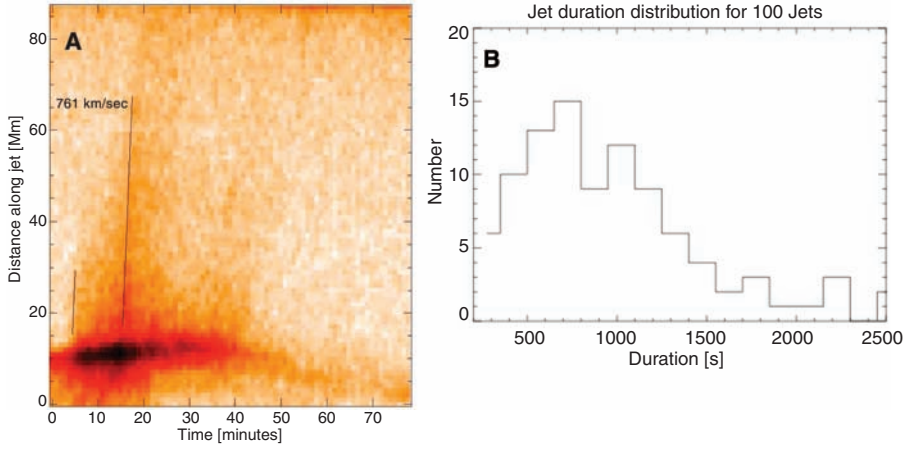


Fig. 3. (**A**) A time-distance plot for the jet shown in Fig. 2. (**B**) Histogram showing the distribution in lifetimes for 100 jets.

after reconnection. We have also found that this high-speed mass flow sometimes occurs multiple times per jet; presumably, there is an Alfvén wave generated during each burst of reconnection.

Alfvén waves are created when the magnetic field is subjected to transverse motions. In the low-β corona, this occurs during reconnection. These transverse motions could be observed in jets as the plasma is constrained to move with the magnetic field. Previous observations with instruments from the Transition Region and Coronal Explorer, Yohkoh, and other missions lacked the imaging cadence and dynamic range to capture these oscillations. We have observed such transverse oscillations for jets (movie S1) using the Hinode XRT. These oscillations have a period of about 200 s and a peak-to-peak magnitude of 8000 km. These observations of the transverse motion of the jet relative to the jet outflow direction are further evidence for Alfvén wave creation during reconnection.

We also found that the frequency of jet formation is at least an order of magnitude greater than that which has been previously reported (*10*–*12*). These previous studies found only a few events per day, primarily resulting from reduced cadence and longer exposures. Our Hinode XRT observations indicate an average of 10 jet events per hour during 100 hours of observations (movie S2). We found that jets frequently occur from the same x-ray bright points or from bright points that formed extremely close to the locations of previous jet-initiation sites, whereas other jets occur from transient x-ray bright points. The x-ray jets in our observations have lifetimes ranging from 100 to 1500 s, much longer than the typical 10- to 100-s lifetimes of the chromospheric jets (*13*, *14*). The histogram in Fig. 3B shows the distribution of jet lifetimes for 100 events observed during the study. No correlation is yet available regarding the location of the bright points, jet formation, and distance from the coronal hole boundary.

Given the increased number of jets observed, we have performed a simple calculation of the impact on the mass loading to the fast solar wind. Each event has an average m_p loss of 1×10^{37}. For both north and south polar coronal holes, XRT recorded, on average, 10 jets per hour. This produces a net flux of 1×10^{12} protons m^{-2} s^{-1} at 1 astronomical unit. Current estimates of the average solar wind flux are only a factor of 10 more than the value of this jet mass-loading contribution (*5*, *10*).

XRT has shown that there is a similarity between these small-scale eruptive events (jets) and the flares/CMEs from active regions, in the sense that both phenomena may involve topological changes to the coronal magnetic field via reconnection, thus providing the energy for both heating and accelerating coronal plasma. This correspondence between events that differ in energy by three to four orders of magnitude may serve as a prototype for the formation of other astrophysical jets, where the energy released is many orders of magnitude greater still.

References and Notes

1. M. Shimojo *et al.*, *Publ. Astron. Soc. Jpn.* **48**, 123 (1996).
2. K. Shibata *et al.*, *Publ. Astron. Soc. Jpn.* **44**, L173 (1992).
3. E. N. Parker, *Astrophys. J.* **128**, 664 (1958).
4. S. Chapman, *Smithson. Contrib. Astrophys.* **2**, 1 (1957).
5. S. R. Cranmer, G. B. Field, J. L. Kohl, *Astrophys. J.* **518**, 937 (1999).
6. S. R. Cranmer, A. A. van Ballegooijen, *Astrophys. J. Suppl. Ser.* **156**, 265 (2005).
7. S. R. Cranmer, A. A. van Ballegooijen, R. J. Edgar, *Astrophys. J. Suppl. Ser.* **171**, 520 (2007).
8. W. H. Matthaeus, G. P. Zank, S. Oughton, D. J. Mullan, P. Dmitruk, *Astrophys. J.* **523**, L93 (1999).
9. M. Velli, *Astron. Astrophys.* **308**, 228 (1993).
10. Y.-M. Wang, *Astrophys. J.* **410**, L123 (1993).
11. B. E. Wood, M. Karovska, J. W. Cook, R. A. Howard, G. E. Brueckner, *Astrophys. J.* **523**, 444 (1999).
12. D. Dobrzycka, S. R. Cranmer, J. C. Raymond, D. A. Biesecker, J. B. Gurman, *Astrophys. J.* **565**, 621 (2002).
13. K. P. Dere, J.-D. F. Bartoe, G. E. Brueckner, *Astrophys. J.* **267**, L65 (1983).
14. B. Schmieder, M. A. Raadu, P. Démoulin, K. P. Dere, *Astron. Astrophys.* **213**, 402 (1989).

15. Hinode is a Japanese mission developed and launched by ISAS/JAXA with NAOJ as a domestic partner and with NASA and the Science and Technology Facilities Council (UK) as international partners. The mission is operated by these agencies in cooperation with the European Space Agency and the Norwegian Space Centre. U.S. members of the XRT team are supported by NASA contract NNM07AA02C to the Smithsonian Astrophysical Observatory.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1580/DC1
SOM Text
References
Movies S1 and S2

26 June 2007; accepted 7 November 2007
10.1126/science.1147050

REPORT

Fine Thermal Structure of a Coronal Active Region

Fabio Reale,[1,2]* Susanna Parenti,[3] Kathy K. Reeves,[4] Mark Weber,[4] Monica G. Bobra,[4] Marco Barbera,[1,2] Ryouhei Kano,[5] Noriyuki Narukage,[6] Masumi Shimojo,[7] Taro Sakao,[6] Giovanni Peres,[1,2] Leon Golub[4]

The determination of the fine thermal structure of the solar corona is fundamental to constraining the coronal heating mechanisms. The Hinode X-ray Telescope collected images of the solar corona in different passbands, thus providing temperature diagnostics through energy ratios. By combining different filters to optimize the signal-to-noise ratio, we observed a coronal active region in five filters, revealing a highly thermally structured corona: very fine structures in the core of the region and on a larger scale further away. We observed continuous thermal distribution along the coronal loops, as well as entangled structures, and variations of thermal structuring along the line of sight.

The solar corona is highly structured by the solar magnetic field and is extremely hot ($\geq 10^6$ K). It is widely accepted that the origin of the heating of the hot plasma confined in the closed coronal structures (named loops) lies in the magnetic field (*1*). The determination of the fine thermal structure of the corona is crucial to solving the origin of the

coronal heating—for instance, whether frequent small and rapid heating episodes (the so-called nanoflares) (*2*–*4*) are able to heat the solar corona entirely or partly or only negligibly (*5*). The solar active regions are good laboratories to investigate this question, because they are bright and include a wide variety of structures. However, they are complex as the structures interact at all scales and overlap with each other. Detailed thermal information is provided by spectrometers, mostly in the ultraviolet band, with moderate spatial resolution, although sometimes with uncertainties due to limited photon statistics and inversion methods (*6*). Imaging instruments also provide some information with lower thermal resolution (*7*, *8*). The X-ray Telescope (XRT) on board the Hinode satellite enables substantially improved measurements due to its higher resolution and sensitivity and the availability of several filter bands. Here, we investigate the detailed thermal structure of a solar coronal active region, with observations in several wideband x-ray filters from the XRT (*9*).

The XRT is a grazing-incidence soft x-ray telescope, which is sensitive to the emission of plasma in the temperature range $6.1 < \log T < 7.5$. Several temperature passbands are provided by a selection over nine different filters in the x-ray band plus one in the optical G band continuum. These are held in two filter wheels (FW1 and FW2) located close to the focal plane in front of the charge-coupled device (CCD) camera. The latter has a 1 arc sec pixel, and the full width at half maximum of the point spread function is ≈0.8 arc sec (*9*).

We observed active region AR 07993 while it was located close to the Sun's center. The Hinode/XRT has a field of view of 512 arc sec by 512 arc sec. The data were obtained on 12 November 2006. The filters used were Al_poly (F1), C_poly (F2), Be_thin (F3), Be_med (F4), and Al_med (F5), with exposure times of 0.26, 0.36, 1.44, 8.19, and 16.38 s, respectively. The selected data set covers 1 hour, starting at 13:00 UT, and the time interval between one exposure and the next in the same filter is about 5 min (12 images in each filter). This active region was monitored by the XRT throughout its passage from the east to the west limb. It flared several times during this period. However, during the selected hour, no major flare and no substantial rearrangement of the region morphology occurred (the average variation is ≈ 12%), so that we can average data over the whole hour. We used the current standard XRT software to perform calibration, including corrections for the read-out signal, flat-field, and CCD bias. The observation images were co-aligned with a cross-correlation technique.

The XRT image of the active region (Fig. 1) shows several interesting features: a main toroidal loop arcade (~100 arc sec wide, labeled A, B, and C) surrounding the central black magnetic pole, open to the south, large-scale loops (up to more than 350,000 km long, region L) in the southwest and north regions, a bright nutshell-like loop system south of the magnetic pole [$(X,Y) \approx (270,220)$ arc sec, region K], and an upside-down V-like loop system immediately to the right of the bright nutshell [$(X,Y) \approx (300,240)$ arc sec, region V]. The brightest part of the region is the loop system in the center of the image [$(X,Y) \approx (220,230)$ arc sec, region X]. The image clearly shows a high degree of loop substructuring, and most loops intersect along the line of sight.

Each XRT filter has a different response to plasma temperature, described by a function $G_i(T)$, where i indicates the filter order number. Softer filters F1 and F2 are more sensitive to the emission of cooler plasma than are the harder filters F4 and F5. Images of the same region in different filterbands provide information about the temperature of the emitting plasma. If the emitting plasma is isothermal and optically thin along the line of sight, the ratio between signals in hard and soft filterbands is a function of the temperature only. We devised a filter ratio that takes full advantage of the XRT capabilities, by combining the available filters (combined filter ratio, CFR). We use the ratio between the geometric mean of the emission detected in all the available filters and the emission I_j in a single (j-th) filter given by Eq. 1:

$$CRF_j(T) = \frac{\left(\prod_{i=1}^{n} I_i\right)^{1/n}}{I_j} = \frac{\left(\prod_{i=1}^{n} G_i(T)\right)^{1/n}}{G_j(T)} \quad (1)$$

The geometric mean allows us to normalize the sensitivity of the soft and hard filters, and to optimize the statistics of the regions where the signal is high in most filters. A further improvement is obtained with the product of the ratios computed for the two softest filters, i.e., Al_poly and C_poly:

$$CIFR(T) = CFR_1(T) \times CFR_2(T) \quad (2)$$

The resulting map of this combined improved filter ratio (CIFR) (Fig. 2B) shows the very high degree of thermal structuring in the core loop arcade. The thickness of the structures is determined down to the order of the instrument resolution (~2 pixels). As seen on the right side of the arcade (region C in Fig. 1), the loops are disposed in an ordered sequence with small overlap. A continuous thermal distribution can be clearly traced along each of these loops. On the left side (region A), there is a much higher level of overlap and even tangling of the thin loop components. This is the region that flared several times during the entire period of XRT observation. The bright nutshell-like loop system (K) in the emission image is clearly linked to the left side of the arcade and is bounded by a very thin bright ring. The V-like loop system (V) is very bright, indicating hot confined plasma. Although still clearly structured, the thermal distribution appears less organized in the region on the left of arcade A. In general, the hot loops that are entirely visible exhibit a hot (bright) apex and a cooling trend along both legs, as expected from standard loop models (*10*, *11*).

[1]Dipartimento di Scienze Fisiche ed Astronomiche, Sezione di Astronomia, Università di Palermo, Piazza del Parlamento 1, 90134 Palermo, Italy. [2]Istituto Nazionale di AstroFisica/Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, 90134 Palermo, Italy. [3]Royal Observatory of Belgium, 3 Circular Avenue, B-1180 Brussels, Belgium. [4]Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA. [5]National Astronomical Observatory, Mitaka, Tokyo 181-8588, Japan. [6]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Sagamihara, Kanagawa 229-8510, Japan. [7]Nobeyama Solar Radio Observatory, National Astronomical Observatory, Nobeyama, Nagano 384-1305, Japan.

*To whom correspondence should be addressed. E-mail: reale@astropa.unipa.it

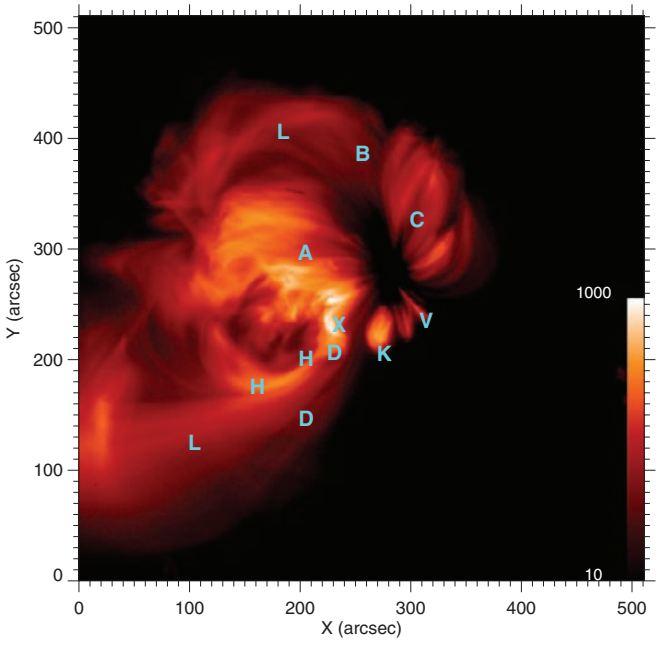


Fig. 1. Emission image of the active region AR 07993 observed with the Hinode/XRT. The image was computed as the geometric mean of the time-averaged (1 hour) images obtained in each filter. The intensity (red) color scale is logarithmic in the range of digital numbers rate (DN/s) indicated in the palette. The figure cuts off the signal below 10 DN pixel^{-1} s^{-1}. Labeled regions are discussed in the text.

In region L on the bottom left of Fig. 1, the largest-scale structures are also mostly on the brightest side of the color scale. Some substructuring is resolved, but on larger scales; the granularity in this region is related to the lower count statistics and does not permit us at present to determine whether the lower level of substructuring is real or due to low statistics.

Assuming an isothermal plasma along the line of sight, the temperature range of this active region plasma would be $6.15 \leq \log T \leq 6.6$ (Fig. 2). If the plasma were not isothermal along the line of sight, the temperature should be taken as an average. Nevertheless, one may evaluate the effect of variations of the emission measure distribution along the line of sight on the value of the filter ratio (Fig. 2). In particular, an excess of hot plasma along the line of sight leads to a notable increase (~0.02 for the selected example in Fig. 2) in the ratio value with respect to the isothermal one, and an excess of cool plasma to a decrease (~0.005 to 0.01).

We realize (Fig. 2) that in some regions there are clear indications of broader emission measure distributions, whereas in other regions the indications are less clear. Bright excesses are associated with the overlap of different large loops [e.g., regions H in Fig. 1, $(X,Y) \sim (200,200)$ arc sec, $(X,Y) \sim (150,170)$ arc sec] and the thin strands in the core region (A). This can be immediately recognized in the thermal map from the steep unexpected change of color (brightening) along the loops, with no appreciable correspondence in the emission map (Fig. 1). Relative dimmings (e.g., regions D in Fig. 1, $(X,Y) \sim (230,200)$ arc sec, $(X,Y) \sim (200,140)$ arc sec] clearly break otherwise smooth ratio distributions along the large loop systems. Such dimmings are also present in the nutshell structure (K), on the right side of the loop arcade [C, $(X,Y) \sim (320,300)$ arc sec], and especially at

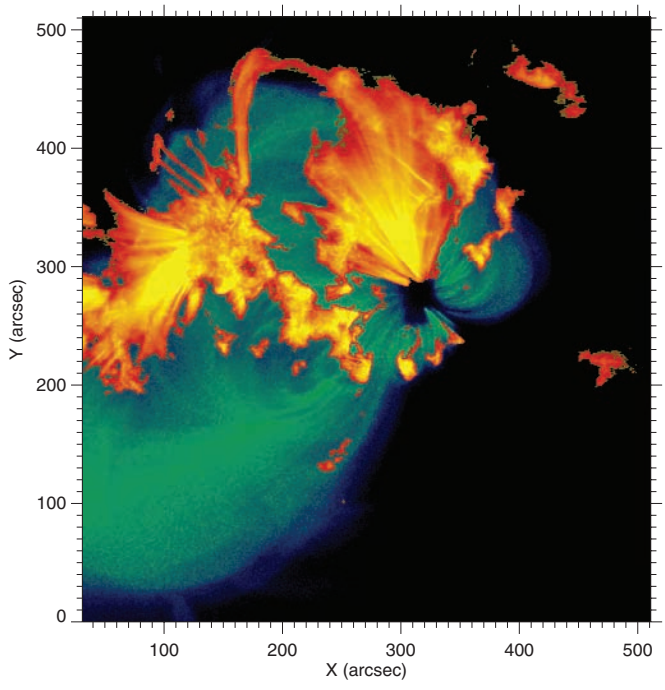


Fig. 3. Same as Fig. 2 (green color scale) superimposed with an emission map taken simultaneously with the TRACE telescope in the 171 A filter passband (sensitive to ~ 1 MK plasma). Only the TRACE emission above 1.5 DN pixel^{-1} s^{-1} is shown (in a yellow color scale).

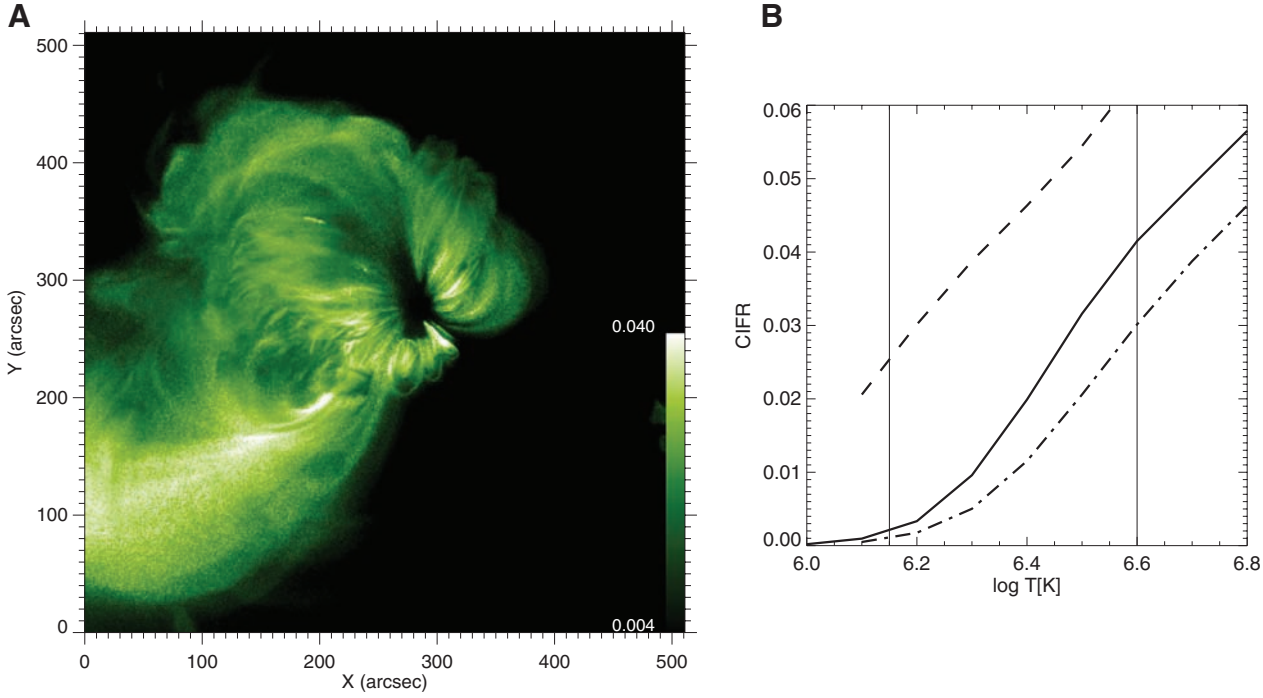


Fig. 2. (**A**) Map of improved combined filter ratio (CIFR) of the active region in Fig. 1. All pixels with low signal (≤ 100 DN/s) in the two softest filters (i.e., Al_poly and C_poly) have been excluded. The green scale of the CIFR is linear in the range indicated in the palette. The CIFR versus temperature correspondence is plotted in (**B**), as predicted [with a coronal plasma emission model, ATOMDB/APEC (*9*)] for an isothermal plasma (thick solid line) and for broader emission measure distributions along the line of sight, i.e. with hot excess (flat emission measure extended to $\Delta\log T = +0.45$, dashed line) and cool excess (flat emission measure extended to $\Delta\log T = -0.45$, dotted-dashed line). The approximate range of the color scale of the CIFR map is indicated between the vertical lines.

the base of the north part of the arcade (region B). Therefore, we can distinguish directly regions where the emission measure distribution along the line of sight is broader and whether the plasma excess is cool or hot.

Superimposing the CIFR map with an emission map in the TRACE 171 A filter, which is sensitive mostly to cooler plasma around 1 MK, we see that this scenario is reliable (Fig. 3): The dimmed spots (D and B) clearly correspond to bright regions in the TRACE image, i.e., there is an excess of cool plasma along the line of sight.

The CIFR map (Fig. 2) shows directly the existence of thermally coherent thin magnetic structures and that hot coronal loops are highly transversally structured. According to theoretical conjectures (*5*, *12*), the tangling of magnetic flux tubes is a prerequisite for magnetic energy release. In this perspective, the tightness and entwining of the thin thermal structures on the left side of the core loop arcade are a direct indication that this region is the probable site of frequent major heating episodes.

References and Notes

1. G. S. Vaiana, A. S. Krieger, A. F. Timothy, *Solar Phys.* **32**, 81 (1973).
2. E. N. Parker, *Astrophys. J.* **330**, 474 (1988).
3. P. J. Cargill, *Astrophys. J.* **422**, 381 (1994).
4. P. J. Cargill, J. A. Klimchuk, *Astrophys. J.* **605**, 911 (2004).
5. J. A. Klimchuk, *Solar Phys.* **234**, 41 (2006).
6. K. P. Dere, H. E. Mason, *Solar Phys.* **144**, 217 (1993).
7. S. Tsuneta *et al.*, *Publ. Astron. Soc. Jpn.* **44**, L63 (1992).
8. T. Yoshida, S. Tsuneta, *Astrophys. J.* **459**, 342 (1996).
9. L. Golub *et al.*, *Solar Phys.* **243**, 63 (2007).
10. J. F. Vesecky, S. K. Antiochos, J. H. Underwood, *Astrophys. J.* **233**, 987 (1979).
11. S. Serio, G. Peres, G. S. Vaiana, L. Golub, R. Rosner, *Astrophys. J.* **243**, 288 (1981).
12. M. C. López Fuentes, J. A. Klimchuk, P. Démoulin, *Astrophys. J.* **639**, 459 (2006).

13. Hinode is a Japanese mission developed and launched by the Institute of Space and Astronautical Science/Japan Aerospace Exploration Agency, with the National Astronomical Observatory of Japan as domestic partner and NASA and Science and Technology Facilities Council (UK) as international partners. It is operated by these agencies in cooperation with the European Space Agency (ESA) and the Norwegian Space Centre (Norway).

F.R., G.P., and M.B. acknowledge support from Italian Ministero dell'Università e Ricerca and Agenzia Spaziale Italiana (ASI), under contracts I/035/05/0 and I/015/07/0. U. S. members of the XRT team are supported by NASA contract NNM07AA02C (to the Smithsonian Astrophysical Observatory). S.P. acknowledges support from the Belgian Federal Science Policy Office through the ESA–Programme de Développement d'Experiences scientifiques program. This work was partially supported by the International Space Science Institute in the framework of an international working team.

18 June 2007; accepted 8 November 2007
10.1126/science.1146590

REPORT

Continuous Plasma Outflows from the Edge of a Solar Active Region as a Possible Source of Solar Wind

Taro Sakao,[1]* Ryouhei Kano,[2] Noriyuki Narukage,[1] Jun'ichi Kotoku,[2] Takamasa Bando,[2] Edward E. DeLuca,[3] Loraine L. Lundquist,[3] Saku Tsuneta,[2] Louise K. Harra,[4] Yukio Katsukawa,[2] Masahito Kubo,[5] Hirohisa Hara,[2] Keiichi Matsuzaki,[1] Masumi Shimojo,[6] Jay A. Bookbinder,[3] Leon Golub,[3] Kelly E. Korreck,[3] Yingna Su,[3] Kiyoto Shibasaki,[6] Toshifumi Shimizu,[1] Ichiro Nakatani[1]

The Sun continuously expels a huge amount of ionized material into interplanetary space as the solar wind. Despite its influence on the heliospheric environment, the origin of the solar wind has yet to be well identified. In this paper, we report Hinode X-ray Telescope observations of a solar active region. At the edge of the active region, located adjacent to a coronal hole, a pattern of continuous outflow of soft-x-ray–emitting plasmas was identified emanating along apparently open magnetic field lines and into the upper corona. Estimates of temperature and density for the outflowing plasmas suggest a mass loss rate that amounts to ~1/4 of the total mass loss rate of the solar wind. These outflows may be indicative of one of the solar wind sources at the Sun.

Since early studies on comet tail orientations in the middle of the twentieth century (*1*, *2*), it has been widely perceived that the interplanetary space around the Sun, the heliosphere, is permeated with continual but varying flows of charged particles: This material flow is the solar wind (*3*, *4*). A huge amount of material is expelled from the Sun in the solar wind, at a rate reaching as high as 1×10^{12} g s^{-1}, into the heliosphere. It extends beyond Jupiter, showing a variety of interactions with planets, including Earth. Observations with space-borne instruments have so far revealed various basic features of the solar wind, such as the existence of two distinct categories of wind velocities: one, the fast solar wind, with velocity as high as ~800 km s^{-1}, and the other, the slow wind, with velocity ~300 km s^{-1}. Nonetheless, at least two fundamental issues remain unresolved: One is the location of the source of the outflows (i.e., the source regions) on the Sun, and the other is the acceleration mechanism of the flows after departing the solar surface.

It is now well recognized that the fast solar wind originates from (polar) coronal holes (*5*, *6*). Recent observations with the Solar and Heliospheric Observatory (SOHO) satellite have enabled us to investigate polar coronal holes with line-of-sight Doppler measurements (*7*, *8*). However, imaging observations of the outflowing material are still rare. The situation is even more uncertain for the case of the slow solar wind, in which multiple source regions have been postulated, such as the boundary of polar coronal holes, helmet streamer structures (from near the top of closed loop structures in the solar corona) (*9*, *10*), or from the edge of active regions (*11*). But again, no clear identification of the source outflow has so far been made.

In this paper, we report imaging observations of a solar active region made with the X-ray Telescope (XRT) (*12*) aboard the Hinode (*13*) satellite. XRT is a grazing incidence soft-x-ray imager that achieves high angular resolution (consistent with 1 arcsec pixel size) with broad and continuous temperature coverage for coronal plasmas ranging from 1 million kelvin (MK) to 10 MK. The XRT observations presented here reveal a pattern of continuous outflow of soft-x-ray–emitting plasmas along apparently open field lines from the edge of an active region butting up adjacent to a coronal hole. Furthermore, the observations from XRT are strongly supported by observations with the Extreme Ultraviolet (EUV) Imaging Spectrometer (EIS) (*14*), also aboard Hinode, which detected persistent upward Doppler motions of coronal plasmas from the apparent outflowing region throughout the XRT observation interval. We assert that these observations are possibly the first identification of outflowing solar wind material.

XRT observed the active region NOAA AR 10942 for 3 consecutive days, from 20 to 22 February 2007, over which period the region was

[1]Institute of Space and Astronautical Science (ISAS), Japan Aerospace Exploration Agency (JAXA), 3-1-1 Yoshinodai, Sagamihara, Kanagawa 229–8510, Japan. [2]National Astronomical Observatory of Japan (NAOJ), 2-21-1 Osawa, Mitaka, Tokyo 181–8588, Japan. [3]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street MS 58, Cambridge, MA 02138, USA. [4]University College London (UCL)–Mullard Space Science Laboratory, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK. [5]High Altitude Observatory, National Center for Atmospheric Research (NCAR), Post Office Box 3000, Boulder, CO 80307–3000, USA. [6]Nobeyama Radio Observatory, NAOJ, 462-2 Nobeyama, Minamimaki, Minamisaku, Nagano 384–1305, Japan.

*To whom correspondence should be addressed. E-mail: sakao@solar.isas.jaxa.jp

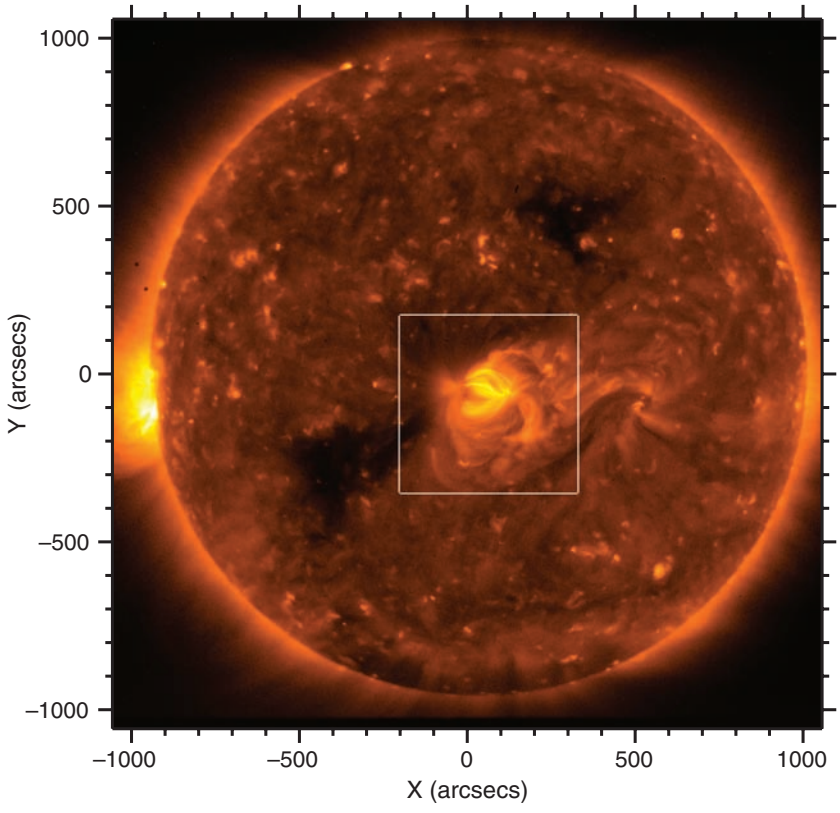


Fig. 1. A full-Sun image taken with the Ti-poly filter of the XRT at 11:13:45 UT on 22 February 2007, shown with a logarithmic intensity scale. Solar north is up and east to the left in this and in all solar images in this paper. Exposure duration is 8 s. The square box indicates the area shown in Fig. 2 and movie S1. The box location on the Sun moves westward with time while the spacecraft pointing tracks solar rotation.

located east of solar disk center near the equator. The observations were chiefly made with XRT's titanium-on-polyimide (Ti-poly) filter, with the size of the readout area on the charge-coupled device (CCD) being 512 pixels by 512 pixels, which was sufficient to cover the entire active region and the surrounding corona. The focus position of the telescope was set to provide the highest angular resolution across the region. XRT also made a series of full-Sun exposures about every 6 hours with Ti-poly and thin-aluminum–on–mesh (thin-Al–mesh) filter pairs, with which fainter (nonactive region) coronal features can be well imaged. The active region was located just west of a large coronal hole (Fig. 1). The coronal hole was long-lived, being present at least one solar rotation both before and after the observation period.

A prominent feature seen in the XRT images (movie S1) is a pattern of plasma outflow along a bundle of fanlike magnetic field lines emanating from a region located at the east edge of the active region. At least the left (east) half of the field line bundle (Fig. 1) is likely to be opened, that is, extending into the outer corona. This outflowing feature was continuously present throughout the entire 3-day observation period. In an attempt to estimate transverse velocity (velocity in the plane orthogonal to the line-of-sight direction) of the outflow, we created a time-distance plot from the Ti-poly images. The time-distance plot in Fig. 2 shows patterns in the flow with transverse velocity ranging between 100 to 170 km s^{-1}, with a typical value of 140 km s^{-1}.

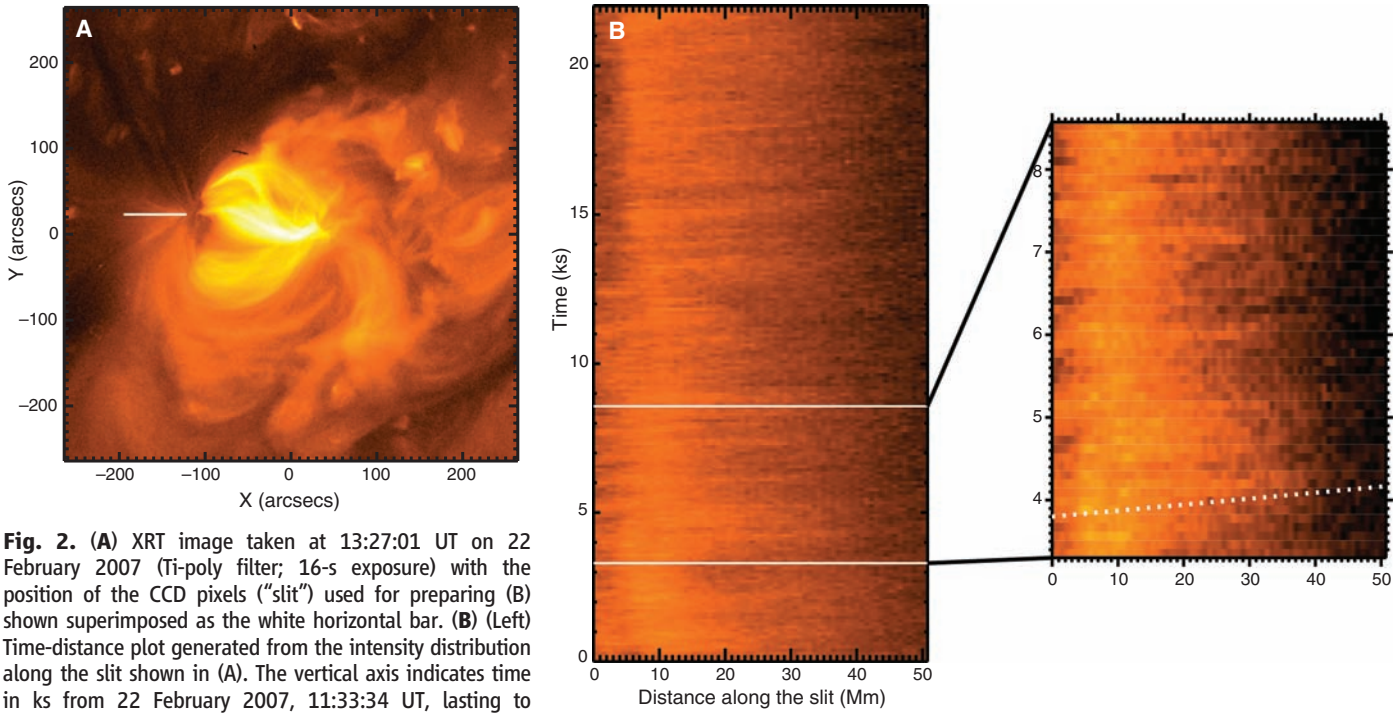


Fig. 2. (**A**) XRT image taken at 13:27:01 UT on 22 February 2007 (Ti-poly filter; 16-s exposure) with the position of the CCD pixels ("slit") used for preparing (B) shown superimposed as the white horizontal bar. (**B**) (Left) Time-distance plot generated from the intensity distribution along the slit shown in (A). The vertical axis indicates time in ks from 22 February 2007, 11:33:34 UT, lasting to 17:40 UT. The horizontal axis gives distance along the slit in Mm measured from the west (right) edge of the slit toward east (left). (Right) Expanded display for the interval indicated by the two white lines in the left image (12:28:36 to 13:56:28 UT). The dotted line represents west-to-east transverse velocity of 140 km s^{-1}.

Whereas the observation was mostly made with the Ti-poly filter for the 3 days, imagery of the region with the thin-Al–mesh filter was also made on 21 February (10:42 to 23:59 UT). By using pairs of images taken with both filters, we can determine an estimate of the physical properties of the plasma with the filter-ratio method (*15*). For the source region, we estimated the temperature ~1.1 MK and emission measure ~7.9×10^{42} cm^{-3} for a volume that corresponds to a single CCD pixel times line-of-sight depth (Fig. 3). Assuming a line-of-sight depth of 1.5 Mm, this gives a density of the region of ~3.2×10^{9} cm^{-3}. With the typical transverse outflow velocity of 140 km s^{-1}, a crude estimate on the amount of the outflowing material per unit time is then ~2.8×10^{11} g s^{-1}. If all the outflowing material were to escape to the interplanetary space along open field lines, this would be equivalent to ~1/4 of the total mass loss rate for the solar wind.

The source region of the outflows, which corresponds to an ensemble of small sunspots (or pores) as seen by the Solar Optical Telescope (SOT) (*16*, *17*) aboard Hinode, does not show any particular brightening activities in soft x-rays associated with the outflows. This is also the case throughout the entire observation period.

In the XRT images, the fanlike field lines along which materials show outflowing motion do not have obvious corresponding conjugate footpoints on the east side. This, however, does not immediately imply that the field lines open out into interplanetary space, because it is possible that these fields form large-scale closed loops but only a portion of such loops is visible in soft x-rays because of temperature and/or density distributions along the loops. On the other hand, computations show (Fig. 4) that potential field lines from the source region at the active region–coronal hole boundary reach the source surface, indicating open field lines. This suggests that at least a fraction of the outflowing material along the fanlike field lines escapes into interplanetary space, resulting in the solar wind.

It has been reported, on the basis of observations with the Transition Region and Coronal Explorer (TRACE) observatory (*18*)

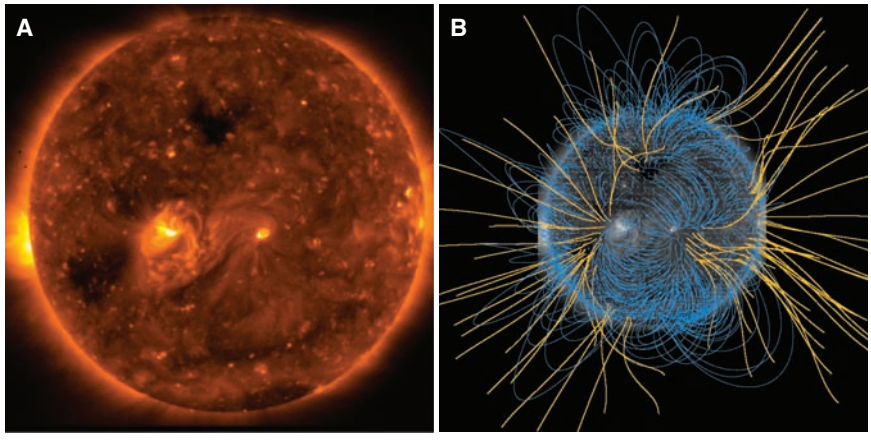


Fig. 4. (**A**) XRT full-Sun image displayed with a logarithmic intensity scale at 00:02:13 UT on 21 February 2007, taken with the thin-Al–mesh filter (4-s exposure). The active region NOAA AR 10942 is slightly to the east of Sun center. (**B**) Potential field lines calculated from a magnetogram taken by SOHO Michelson Doppler Imager (MDI) at 00:03:02 UT on 21 February 2007, with the source surface set at 2.5 solar radii. Field lines reaching the source surface, i.e., those considered to open out into interplanetary space, are shown in yellow, whereas closed lines are shown in blue. The field lines are overlaid on the same XRT image as that in (A). Because of lack of magnetogram data for the polar regions, field lines emanating from the polar regions should not be regarded as valid (such regions were filled with a null magnetic signal when performing the potential field calculation).

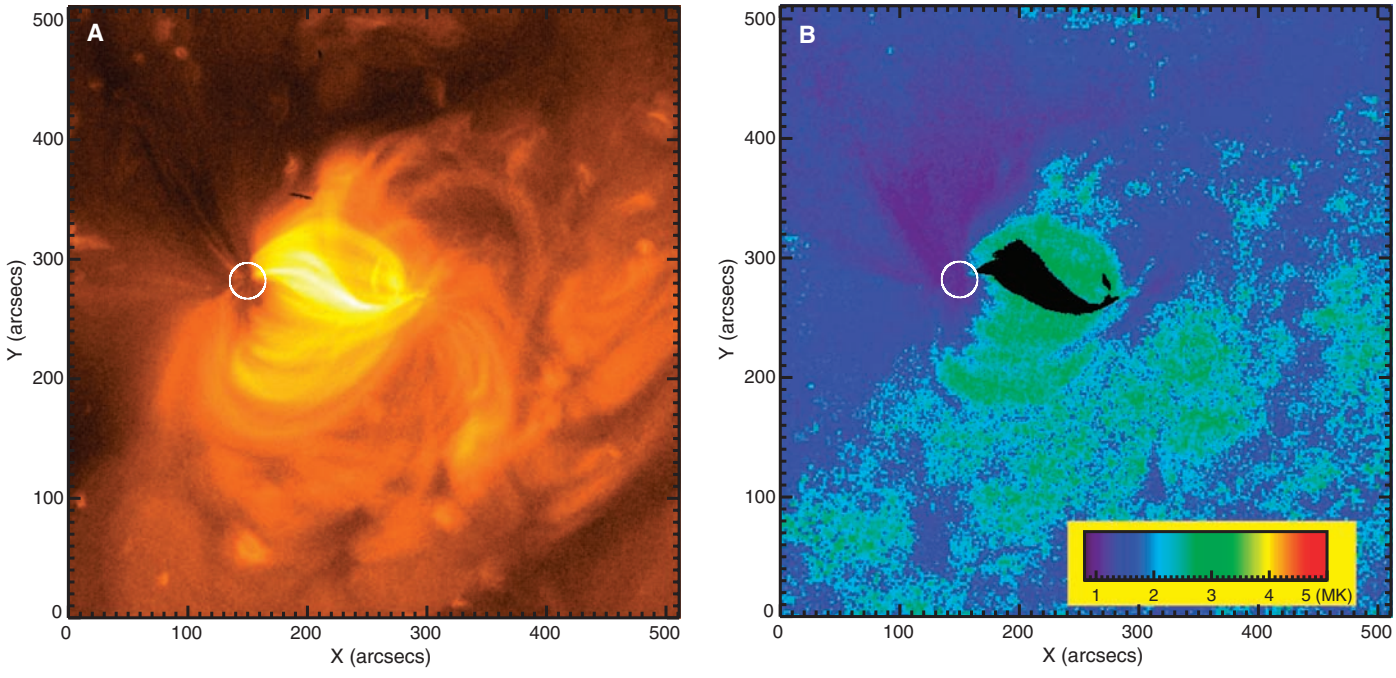


Fig. 3. (**A**) XRT image of NOAA AR 10942 taken at 23:57:45 UT on 21 February 2007 with the Ti-poly filter. The source region of the outflows is shown by the circle, whose radius is 10.9 Mm. (**B**) Map of filter-ratio temperatures for the same area as in (A), derived from a pair of images taken with the thin-Al–mesh and Ti-poly filters at 23:57:14 and 23:57:45 UT, respectively, on 21 February 2007. Exposure duration for each image is 16 s. Black areas in the active region correspond to saturated CCD pixels in the thin-Al–mesh image, where filter-ratio temperatures could not be derived with this pair of images. The white circle indicates the same region as (A).

and the EUV Imaging Telescope (EIT) (*19*) aboard SOHO, that traveling disturbances exist along fanlike coronal magnetic field lines and that these have been identified as slow magnetoacoustic waves with period of 3 min (*20*, *21*). Meanwhile, it has been argued that there are flows along coronal loops as observed by TRACE and SOHO, with velocities of several tens of km s^{-1} in emission lines at temperatures around 0.6 to 1 MK (*22*, *23*). Although the relationship between the flow pattern reported in this study and the wave phenomena seen with TRACE and EIT is a subject for subsequent studies, EIS aboard Hinode, with which line-of-sight Doppler velocities of coronal plasmas can be measured, also identified upward Doppler velocities of ~50 km s^{-1} in a coronal emission line (Fe XII) at the outflowing region in each of the 3 days of the XRT observations (EIS data from 23:45 UT on 20 February, 05:32 UT on 21 February, and 18:08 UT on 22 February, with the last one indicating velocity possibly as high as ~90 km s^{-1}). These Doppler velocities may be consistent with the typical transverse velocity of 140 km s^{-1} obtained from XRT, considering the inclination of the field lines. The EIS upward Doppler signatures add support to the interpretation of the XRT observations as evidence for the presence of outflowing plasmas.

The assertion that the observed outflows are a possible source of the solar wind is also supported by interplanetary scintillation tomographic observations of the solar wind (*11*), which argue that the low-speed solar wind is most likely associated with regions that include rapidly expanding open magnetic flux adjacent to active regions. Furthermore, in situ measurements of solar wind particles and magnetic field with the Advanced Composition Explorer (ACE) satellite support the idea that one of the sources of the slow solar wind resides in boundaries between coronal holes and active regions (*24*). The striking resemblance in the magnetic field configuration proposed in those studies and the present observations suggest that the observed outflows correspond to one of the sources of the slow solar wind.

References and Notes

1. C. Hoffmeister, *Z. Phys.* **22**, 265 (1943).
2. L. Biermann, *Z. Phys.* **29**, 274 (1951).
3. S. Chapman, *Smithson. Contrib. Astrophys.* **2**, 1 (1957).
4. E. N. Parker, *Astrophys. J.* **128**, 664 (1958).
5. G. Noci, *Sol. Phys.* **28**, 403 (1973).
6. B. Bell, G. Noci, *J. Geophys. Res.* **81**, 4508 (1976).
7. D. M. Hassler *et al.*, *Science* **283**, 810 (1999).
8. C.-Y. Tu *et al.*, *Science* **308**, 519 (2005).
9. Y.-M. Wang *et al.*, *Astrophys. J.* **498**, L165 (1998).
10. J. T. Gosling *et al.*, *J. Geophys. Res.* **86**, 5438 (1981).
11. M. Kojima *et al.*, *J. Geophys. Res.* **104**, 16993 (1999).
12. L. Golub *et al.*, *Sol. Phys.* **243**, 63 (2007).
13. T. Kosugi *et al.*, *Sol. Phys.* **243**, 3 (2007).
14. J. L. Culhane *et al.*, *Sol. Phys.* **243**, 19 (2007).
15. H. Hara, S. Tsuneta, J. R. Lemen, L. W. Acton, J. M. McTiernan, *Publ. Astron. Soc. Jpn.* **44**, L135 (1992).
16. K. Ichimoto and the Solar-B Team, *J. Korean Astron. Soc.* **38**, 307 (2005).
17. T. Shimizu, in *The Solar-B Mission and the Forefront of Solar Physics*, T. Sakurai, T. Sekii, Eds., ASP Conference Series vol. 325 (Astronomical Society of the Pacific, San Francisco, CA, 2004), pp. 3–13.
18. B. N. Handy *et al.*, *Sol. Phys.* **187**, 229 (1999).
19. J.-P. Delaboudinière *et al.*, *Sol. Phys.* **162**, 291 (1995).
20. I. De Moortel, J. Ireland, R. W. Walsh, *Astron. Astrophys.* **355**, L23 (2000).
21. E. Robbrecht *et al.*, *Astron. Astrophys.* **370**, 591 (2001).
22. A. R. Winebarger, E. E. DeLuca, L. Golub, *Astrophys. J.* **553**, L81 (2001).
23. A. R. Winebarger, H. Warren, A. van Ballegooijen, E. E. DeLuca, L. Golub, *Astrophys. J.* **567**, L89 (2002).
24. Y.-K. Ko *et al.*, *Astrophys. J.* **646**, 1275 (2006).

25. We are grateful to K. Fujiki of Solar-Terrestrial Environment Laboratory, Nagoya University, and T. K. Suzuki of the University of Tokyo for stimulating discussions on the observed features of the solar wind, the SOHO MDI team for providing full-Sun magnetogram data, and A. C. Sterling of the NASA Marshall Space Flight Center for improvement of the manuscript. Hinode is a Japanese mission developed and launched by ISAS/JAXA, collaborating with NAOJ as a domestic partner and NASA and Science and Technology Facilities Council (STFC) (UK) as international partners. Scientific operation of the Hinode mission is conducted by the Hinode science team organized at ISAS/JAXA. This team mainly consists of scientists from institutes in the partner countries. Support for the postlaunch operation is provided by JAXA and NAOJ (Japan), STFC (UK), NASA (USA), European Space Agency, and Norwegian Space Centre (Norway). U.S. members of the XRT team are supported by NASA contract NNM07AA02C to the Smithsonian Astrophysical Observatory. Lastly, we would like to express our sincere gratitude to the late T. Kosugi, former project manager of the Hinode mission at ISAS/JAXA, whose years of outstanding effort led to the success of this mission.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1585/DC1
Movie S1

2 July 2007; accepted 2 November 2007
10.1126/science.1147292

REPORT

Slipping Magnetic Reconnection in Coronal Loops

Guillaume Aulanier,[1]* Leon Golub,[2] Edward E. DeLuca,[2] Jonathan W. Cirtain,[2] Ryouhei Kano,[3] Loraine L. Lundquist,[2] Noriyuki Narukage,[4] Taro Sakao,[5] Mark A. Weber[2]

Magnetic reconnection of solar coronal loops is the main process that causes solar flares and possibly coronal heating. In the standard model, magnetic field lines break and reconnect instantaneously at places where the field mapping is discontinuous. However, another mode may operate where the magnetic field mapping is continuous but shows steep gradients: The field lines may slip across each other. Soft x-ray observations of fast bidirectional motions of coronal loops, observed by the Hinode spacecraft, support the existence of this slipping magnetic reconnection regime in the Sun's corona. This basic process should be considered when interpreting reconnection, both on the Sun and in laboratory-based plasma experiments.

Magnetic reconnection is a diffusive magnetohydrodynamic (MHD) process, through which magnetic loops exchange their connections at large scales (*1*). In magnetic loops whose footpoints are both rooted in a dense conducting layer, which is the case in the solar atmosphere (*2*, *3*), reconnection also redistributes field-aligned electric currents far from where they were located early on (*4*–*6*). When reconnection occurs across layers where the magnetic mapping is discontinuous (which naturally results from existing magnetic null points), magnetic loops instantaneously exchange their large-scale connections by pairs. This is always true, regardless of the rate of magnetic flux being transferred through the local reconnection site. It is the standard reconnection model, and it is used to model solar flares in general. However, solar observations, combined with theoretical reconstructions of the coronal magnetic field, show that brightenings related to solar flares are often not associated with mapping discontinuities but rather with so-called quasi-separatrix layers (QSLs), across which the magnetic mapping has very sharp spatial gradients, although still being continuous (*7*–*9*). Whether magnetic reconnection, instead of simple magnetic diffusion, occurs in QSLs, and, if so, what are its large-scale consequences, have been debated for more than 10 years, because the absence of mapping discontinuities in QSLs forbids the occurrence of standard reconnection.

Three-dimensional (3D) MHD simulations that included collisional diffusive terms in Ohm's

[1]Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique (LESIA), Observatoire de Paris, Centre National de la Recherche Scientifique (CNRS), Université Pierre et Marie Curie (UPMC), Université Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France. [2]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA. [3]National Astronomical Observatory of Japan (NAOJ), Mitaka, Tokyo 181–8588, Japan. [4]Kwasan and Hida Observatories, Kyoto University, Yamashina, Kyoto 607–8471, Japan. [5]Institute of Space and Astronautical Science (ISAS), Japan Aerospace Exploration Agency (JAXA), 3-1-1 Yoshinodai, Sagamihara, Kanagawa 229–8510, Japan.

*To whom correspondence should be addressed. E-mail: guillaume.aulanier@obspm.fr

Law (*10*–*12*) recently suggested that magnetic field lines can slip along each other in QSLs, at velocities that can be much faster than any local characteristic propagation speed, such as the sound and the Alfvén speeds (*12*). This peculiar behavior is due to resistively driven derotation of the magnetic field locally in QSL current sheets. This process naturally leads to successive rearrangements of the connections between neighboring field lines and hence to large-scale and continuous apparent field line footpoint motions, along the QSLs only. This is almost the same as when reconnection occurs in mapping discontinuities; however, the change in the footpoint connectivity of the field no longer occurs instantaneously but at finite speed. Slip-running and slipping reconnection regimes were respectively defined as super- and sub-Alfvénic field line fast slippage. Both differ from mere magnetic diffusion, which results from slow and nonlocalized drifts of field line footpoints. It has been argued that fast field line slippage often occurs in the solar corona (*12*) and that such slippage may explain the observed fast motions of hard x-ray sources along the flare brightening in many events observed with the Hard X-ray Telescope onboard Yohkoh (*13*) and with the Ramaty High Energy Solar Spectroscopic Imager (*14*). Up to now, direct observations of slipping magnetic reconnection have never been reported.

In this paper, we describe fast slippage of coronal loops, observed by the X-ray Telescope (XRT) onboard Hinode. We use the titanium-polyimide (Ti-poly) filter, whose time cadence of the order of 1 min and whose broad temperature response for plasmas hotter than 10 million K (MK) (*15*) facilitate the observation of rapidly moving loops during their short conductive cooling time scale, after they have been impulsively heated by reconnection. Analysis of context observations [of the low solar atmosphere in the extreme ultraviolet (EUV) by the Transition Region and Coronal Explorer (TRACE) and of photospheric magnetic field measurements by the Michelson Doppler Imager (MDI) onboard Solar and Heliospheric Observatory (SOHO)] supports the interpretation of the apparent XRT loop motions by a slipping magnetic reconnection process.

On 6 February 2007, a bipolar active region AR10940, embedded in the remnants of an older region, was located near the west solar limb. XRT observed the region in soft x-rays with a pixel resolution of 1″ and a 512 by 512 pixel field of view. Between 13:00 and 18:00 UT, XRT observations were captured with both the Ti-poly and the Al-poly filters, alternatively; each filter recorded data with a sequence of short and long exposures taken at ~ 3- to 6-s intervals, alternating between filters every ~ 2 min. Bright coronal loops were identified in AR10940 (with short exposures), as well as fainter remote loops (with long exposures, which resulted in charge-coupled device (CCD) saturation of the bright loops) and a cooler unresolved coronal emission (with the Al-poly filter, which is sensitive to lower temperatures of about 1 MK). We restrict our analysis to the Ti-poly filter observations, because we are only interested in the loop dynamics. The coronal structure of AR10940 had a long, flat, and faint S-shaped loop that connected it to an eastern "quiet" region (i.e., area with an absence of strong sunspot-related magnetic flux concentration), where an ensemble of brighter C-shaped loops was also located. Apparent displacements of the C-shaped loops occur in the eastern part of the active region (movie S1). These apparent motions resemble those of modeled magnetic field lines in the original slip-running reconnection MHD simulations (*12*).

To better understand the nature of this observed behavior, we also used context observations by (i) the SOHO/MDI magnetograph (*16*), which measured the line-of-sight magnetic field of the full Sun every 96 min with a spatial resolution of 1.97″, and (ii) TRACE (*17*), which observed the west solar limb in the EUV every 45 to 73 s with its 195 Å filter with a 0.5″ spatial resolution. The data are readily coaligned with the solar limb as a reference around 14:30 UT (Fig. 1, A to E). The only substantial features that are visible with all instruments are related to two bright points, located to the southeast of the slipping coronal loops. These features appear as small bright EUV and x-ray loops, located above two small and isolated positive-polarity magnetic flux concentrations (shown in white in Fig. 1C). Because these bright points were remote from the area of loop slippage, they probably played no role in the slippage episodes. Other than at the bright points, the soft x-ray loops and the EUV loops never overlay each other. The S-shaped x-ray loop in particular does not appear in the EUV, and most EUV loops seem to reach higher altitudes, although their apices are not visible. This pattern has already been seen with lower spatial resolutions with the Soft X-ray Telescope (SXT) onboard Yohkoh (*18*–*20*). The overlays among observations by XRT, TRACE, and MDI (Fig. 1, D and E) reveal that the slipping x-ray loops are all rooted in a curved north-south elongated EUV brightening (hereafter referred to as a "ribbon") that links several negative-polarity magnetic flux concentrations. Such EUV ribbons have been attributed to complex magnetic topologies, not only during intense solar flares (*9*) but also above quiet solar regions (*21*). Therefore, they are the key feature that supports the idea that a QSL is present in this region.

On the basis of this interpretation, we propose a schematic representation of the magnetic topology

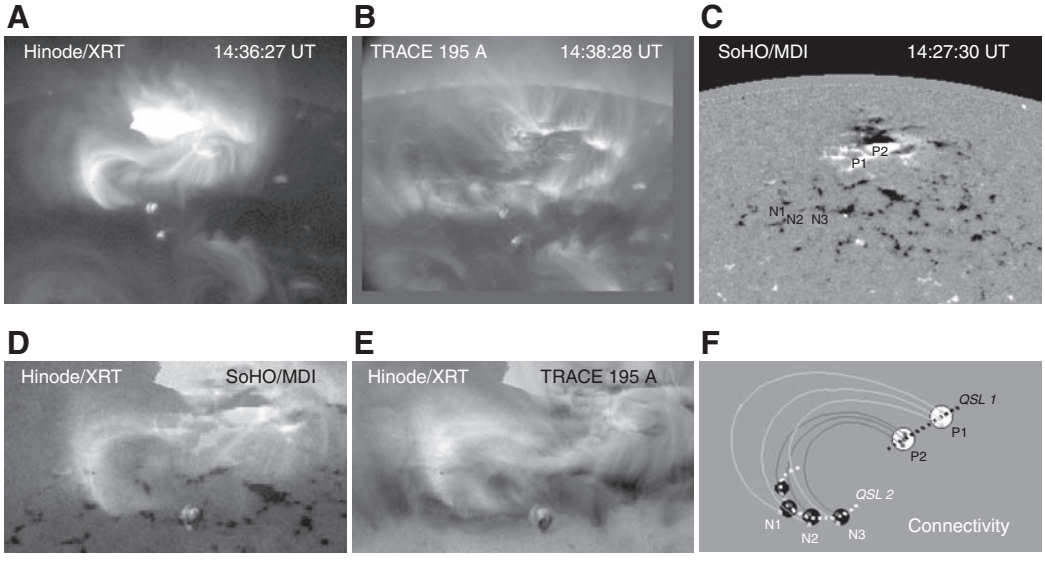


Fig. 1. Coaligned observations of AR10940 near the west limb on 6 February 2007. The images are rotated so that the solar north is to the left. (**A**) Hinode/XRT long-exposure image in soft x-rays with the Ti-poly filter. (**B**) TRACE image in the EUV with the 195 Å filter. (**C**) SOHO/MDI photospheric line-of-sight magnetogram. (**D** and **E**) Superposition of the observations, showing the link between EUV ribbons, soft x-ray coronal loops, and photospheric magnetic fields. (**F**) Schematic representation of the magnetic mapping between the positive polarities of AR10940 (P1 and P2) and the eastern network elements of negative polarity (N1, N2, and N3). The field of view is 512″ by 400″ [for (A) to (C)] and 308″ by 172″ [for (D) and (E)]. A sequence of composites of long- and short-exposure XRT observations and a sequence of TRACE observations are provided as Quicktime animations (movies S1 and S2) in the supporting online material.

of this region (Fig. 1F). In order to calculate the magnetic mapping, we attempted to model the observed plasma features by carrying out a so-called linear force-free reconstruction of the coronal field, using the line-of-sight MDI magnetogram as a boundary condition. This model, however, was not successful. Still, based on the data and using geometrical properties of squashed magnetic flux tubes (*7*, *8*), we conjecture that quasi-separatrices should be present, having one footpoint (QSL2) above the EUV ribbon and another footpoint (QSL1) above the active region positive-polarity strong flux concentrations (P1 and P2). In our sketch (Fig. 1F), the dark and light gray lines represent the C-shaped and S-shaped coronal loops, respectively, observed by XRT. The strong mapping gradients imply that, at a given time (for example, as one selects field line footpoints successively placed across QSL1 within P2), the conjugate field line footpoints quickly and continuously shift from negative-polarity magnetic flux concentrations (N1 to N3). If the flux concentrations were modeled by magnetic charges placed on a photospheric plane, photospheric magnetic null points defining a complex skeleton of mapping discontinuities would replace the QSLs, but, in that case, reconnection would not result in the observed loop slippage.

The XRT time-sequence (movie S1) shows a nearly continuous set of coronal loop slippage events during a 3.5-hour period, between 13:40 and 17:10 UT. Many individual loops display apparent motions both northward and southward. Several loops sometimes move at the same time, either in the same direction or in opposite directions. These apparent motions do not look like loop oscillations. The TRACE time-sequence shows that, during this time interval, the EUV ribbon has a much slower evolution, both in brightness and in position (movie S2). Still, a new bright ribbon oriented in an east-west direction gradually forms around 14:30 UT and slowly separates from the main ribbon. MDI magnetograms also show that the magnetic flux concentrations around and below the EUV ribbon are more or less stationary. We select five clear loop slippage events (Fig. 2) for analysis. We refer to each of these events by a number (event 1 to event 5), which corresponds to the columns in Fig. 2. In all events, the loops slip along the main north-south ribbon, except for those of event 4, which seem to move along the newly formed east-west ribbon. Because of an inhomogeneous and unsteady background of nonslipping loops, we could not highlight the slipping loops on these frames by image treatment. So, we also provide a schematic representation of the slippage of some magnetic loops as observed by XRT (Fig. 3).

The speeds of the footpoint motions were ≈ 130 km/s for event 1, 40 and 80 km/s for event 2, 50 and 150 km/s for event 3, at least 30 km/s for event 4 (projection effects there forbid the estimation of a precise velocity), and 50 km/s for event 5. These speeds are about 10% of the typical Alfvén speeds in the corona, which does not satisfy the slip-running regime sensu stricto. Still, these speeds are of the order of the speed of magnetosonic waves in the solar chromosphere and are two to three orders of magnitude larger than typical photospheric velocities. These velocities also match those observed for some solar flare–related hard x-ray footpoint motions (*13*, *14*).

The non-oscillatory x-ray loop fast slippage in both directions, combined with the occurrence of slowly moving EUV ribbons above nearly stationary photospheric magnetic fields, strongly suggests that the apparent loop motions are neither simple heating wavefronts in the corona nor a result of line-tied photospheric motions or coronal bulk flows. These apparent motions can instead be reasonably interpreted as the signature of the sub-Alfvénic regime of the 3D slip-running reconnection process (*12*). In this context, the observed displacement of x-ray loops corresponds to the gradual brightening of neighboring loops successively heated by reconnection. The finding of this sub-Alfvénic regime is not unexpected in this region, because super-Alfvénic slippage was predicted to occur only in narrow quasi-separatrices, which is not likely the case for the magnetic configuration of this region. It is far from having an overall quadrupolar magnetic field distribution, and the ratio between the magnetic flux of the polarities N1, N2, and N3 and that of the surrounding quiet Sun as measured by MDI is not large. Both conditions must be met in order to have thin quasi-separatrices (*7*, *22*).

Because TRACE did not show the slipping loops for the present event, slipping loops must be heated to higher temperatures than those observable with the current EUV-imaging telescopes. Past soft x-ray observations of this process with Yohkoh/SXT were not reported, but that instrument had a much coarser spatial and temporal resolution than that of XRT. According to MHD theory, slipping reconnection is expected to be very common in the Sun's atmosphere and (more generally) within natural and laboratory

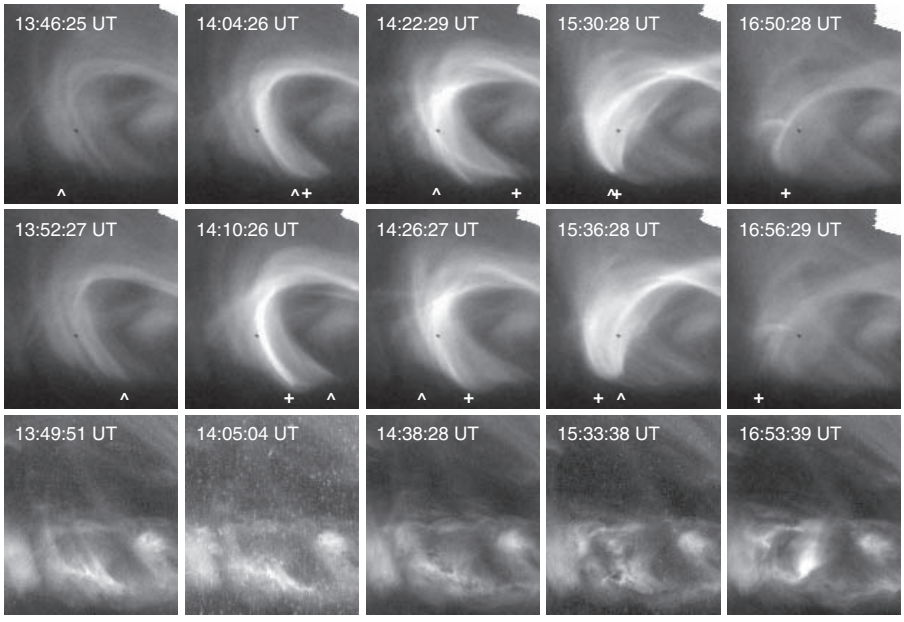


Fig. 2. Selection of five events of coronal loop slippage. Event images are grouped by column, from event 1 (leftmost column) to event 5 (rightmost column). For each event, the first and last images of the moving loops observed by XRT (with the Ti-poly filter) are shown in the top and middle panels, respectively. The evolution of EUV ribbons observed by TRACE (with the 195 Å filter) is shown in the bottom panels. The slipping loop footpoints, indicated by carrot and positive-sign symbols, move only leftward for the 14:22 and 16:50 UT events, only rightward for the 13:46 UT event, and in both directions simultaneously for the 14:04 and 15:30 UT events. The field of view is 121″ by 139″.

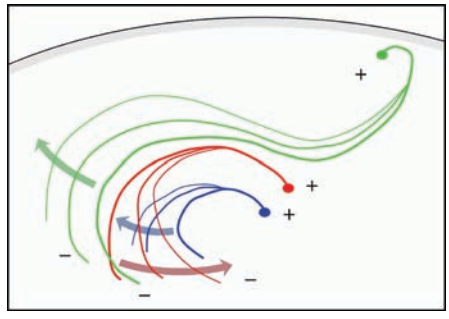


Fig. 3. Schematic representation of the time-evolution of three slipping loops adapted from XRT observations. The loops have fixed footpoints (indicated by circles) rooted in positive-polarity magnetic fields (indicated by positive signs). Their conjugate footpoints slip in the negative-polarity fields (indicated by negative signs). The slippage directions are indicated by the arrows. The solar limb is shown as a gray arc for reference.

plasmas that are confined by 3D magnetic fields, such as in some astrophysical objects and fusion devices after axisymmetry breaks down (for example, as a result of kinking of current-carrying loops). Slipping reconnection theory predicts continuous and fast displacements of impact sites of heat, waves, and energetic particles, originating from a single remote reconnection site. It explains these displacements as being due to the transport of mass and energy along neighboring magnetic field lines, because these lines successively pass through the reconnection site as they continuously change their connectivities.

References and Notes

1. P. A. Sweet, in *Electromagnetic Phenomena in Cosmical Physics*, B. Lehnert, Ed. (Cambridge Univ. Press, Cambridge, 1958), pp. 123–134.
2. B. Leroy, *Astron. Astrophys.* **91**, 136 (1980).
3. L. Ofman, *Astrophys. J.* **568**, L135 (2002).
4. T. Yokoyama, K. Shibata, *Astrophys. J.* **436**, L197 (1994).
5. Z. W. Ma, C. S. Ng, X. Wang, A. Bhattacharjee, *Phys. Plasmas* **2**, 3184 (1995).
6. J. T. Karpen, S. K. Antiochos, C. R. DeVore, L. Golub, *Astrophys. J.* **495**, 491 (1998).
7. P. Démoulin, J.-C. Hénoux, E. R. Priest, C. H. Mandrini, *Astron. Astrophys.* **308**, 643 (1996).
8. V. S. Titov, G. Hornig, P. Démoulin, *J. Geophys. Res.* **107**, SSH3-1 (2002).
9. P. Démoulin, *Adv. Space Res.* **37**, 1269 (2006).
10. D. I. Pontin, K. Galsgaard, G. Hornig, E. R. Priest, *Phys. Plasmas* **12**, 052307 (2005).
11. I. De Moortel, K. Galsgaard, *Astron. Astrophys.* **451**, 1101 (2006).
12. G. Aulanier, E. Pariat, P. Démoulin, C. R. DeVore, *Sol. Phys.* **238**, 347 (2006).
13. L. Fletcher, H. Hudson, *Sol. Phys.* **210**, 307 (2002).
14. S. Krucker, G. J. Hurford, R. P. Lin, *Astrophys. J.* **595**, L103 (2003).
15. L. Golub *et al.*, *Sol. Phys.* **243**, 63 (2007).
16. P. H. Scherrer *et al.*, *Sol. Phys.* **162**, 129 (1995).
17. B. N. Handy *et al.*, *Sol. Phys.* **187**, 229 (1999).
18. S. E. Gibson *et al.*, *Astrophys. J.* **574**, 1021 (2002).
19. B. Schmieder *et al.*, *Astrophys. J.* **601**, 530 (2004).
20. A. R. Winebarger, H. P. Warren, *Astrophys. J.* **626**, 543 (2005).
21. T. E. Berger *et al.*, *Sol. Phys.* **190**, 409 (1999).
22. P. Démoulin, E. R. Priest, *Sol. Phys.* **175**, 123 (1997).

23. Hinode is a Japanese mission that was developed, launched, and operated by ISAS/JAXA, in partnership with NAOJ, the National Aeronautics and Space Administration (NASA), and the Science and Technology Facilities Council (UK). Additional operational support is provided by the European Space Agency and the Norwegian Space Centre. This work is supported by a contract from NASA to the Smithsonian Institution and by a contract from the European Commission to the Observatoire de Paris.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1588/DC1
Movies S1 and S2

6 June 2007; accepted 9 November 2007
10.1126/science.1146143

REPORT

Chromospheric Anemone Jets as Evidence of Ubiquitous Reconnection

Kazunari Shibata,[1]* Tahei Nakamura,[1] Takuma Matsumoto,[1] Kenichi Otsuji,[1] Takenori J. Okamoto,[1,2] Naoto Nishizuka,[1] Tomoko Kawate,[1] Hiroko Watanabe,[1] Shin'ichi Nagata,[1] Satoru UeNo,[1] Reizaburo Kitai,[1] Satoshi Nozawa,[3] Saku Tsuneta,[2] Yoshinori Suematsu,[2] Kiyoshi Ichimoto,[2] Toshifumi Shimizu,[4] Yukio Katsukawa,[2] Theodore D. Tarbell,[5] Thomas E. Berger,[5] Bruce W. Lites,[6] Richard A. Shine,[5] Alan M. Title[5]

The heating of the solar chromosphere and corona is a long-standing puzzle in solar physics. Hinode observations show the ubiquitous presence of chromospheric anemone jets outside sunspots in active regions. They are typically 3 to 7 arc seconds = 2000 to 5000 kilometers long and 0.2 to 0.4 arc second = 150 to 300 kilometers wide, and their velocity is 10 to 20 kilometers per second. These small jets have an inverted Y-shape, similar to the shape of x-ray anemone jets in the corona. These features imply that magnetic reconnection similar to that in the corona is occurring at a much smaller spatial scale throughout the chromosphere and suggest that the heating of the solar chromosphere and corona may be related to small-scale ubiquitous reconnection.

Recent solar space missions such as Yohkoh, SOHO, TRACE, and RHESSI have revealed that the outer atmosphere of the Sun, such as the solar corona and transition region, is much more dynamic than had been thought and contains numerous jets and microflares. The mechanism providing such a highly dynamic corona and transition region is a mystery and may be related to the long-standing puzzle of how the corona is heated to a million K (*1*). It has long been observed with Hα observations from the ground that cool jets, called surges, often occur in the chromosphere near sunspots in association with small flares (*2*). Here, we report the discovery of tiny chromospheric anemone jets with the Solar Optical Telescope (SOT) aboard Hinode (*3*). These jets are much smaller and occur much more frequently than surges and are considered to be indirect evidence of small-scale ubiquitous reconnection in the solar atmosphere as conjectured by Parker (*4*).

Figure 1 shows the Ca II H broadband filter snapshot image of the active region on 17 December 2006 near the west limb, taken with Hinode/SOT (see also movie S1). The dark elongated ellipse seen near the limb is a sunspot. Numerous tiny jets ejected from bright points (arrows) can be seen. There are also numerous jets or jetlike structures throughout this image, some of which are spicules, fibrils, and microjets above sunspot penumbra (*5*). We concentrate on the isolated jets from bright points, which we call Ca jets. Figure 2 shows the time evolution of four typical Ca jets. The footpoints of these jets are not a simple bright point, but show a cusp- or inverted Y-shape. This is the characteristic shape of x-ray jets, known as anemone jets (*6*–*8*).

On the basis of Yohkoh observations (*6*), the anemone shape is formed as a result of magnetic reconnection between an emerging magnetic bipole and a preexisting uniform vertical field. Namely, once reconnection occurs between the emerging bipole and vertical field, the field lines with a polarity opposite to that of the ambient field become connected to the ambient polarity regions, forming a fan shape like that of a sea anemone (*6*). Hence, this kind of structure has been called an anemone. X-ray jets, discovered by Yohkoh, tend to have an anemone shape at their footpoint (*9*), providing indirect evidence of magnetic reconnection as a mechanism for jet formation. Most of the footpoints of x-ray jets correspond to mixed-polarity regions (*10*), supporting a magnetic reconnection mechanism. Two-dimensional magnetohydrodynamic (MHD) numerical simulations of reconnection between emerging flux and uniform coronal field also reproduced well the observed characteristics of x-ray jets (*7*, *8*). In this case, the reconnection occurs in the low corona or in the upper chromosphere, producing high-speed jets. Because the characteristic velocity of a reconnection jet is the local Alfvèn speed, the velocity of the high-speed jet becomes comparable to the coronal Alfvèn speed, or 100 to 1000 km s^{-1} or more, which is comparable to the actual observed velocity of x-ray jets. The plasma in the corona is heated to temperatures ranging from a few million K to about 10 million K. This hot plasma can be observed as microflares and soft x-ray jets (Fig. 3A). At the

[1]Kwasan and Hida Observatories, Kyoto University, Yamashina, Kyoto 607–8471, Japan. [2]National Astronomical Observatory, Mitaka, Tokyo 181–8588, Japan. [3]Department of Science, Ibaraki University, Mito, Ibaraki 310–8512, Japan. [4]Institute of Space and Astronautical Science/Japan Aerospace Exploration Agency, Sagamihara, Kanagawa 229–8510, Japan. [5]Lockheed Martin Solar and Astrophysics Laboratory, B/252, 3251 Hanover Street, Palo Alto, CA 94304, USA. [6]High Altitude Observatory, National Center for Atmospheric Research, Post Office Box 3000, Boulder, CO 80307–3000, USA.

*To whom correspondence should be addressed. E-mail: shibata@kwasan.kyoto-u.ac.jp

same time, cool jets are accelerated if the reconnection occurs in the upper chromosphere where cool (10^4 K) plasma is situated near the reconnection point. These cool jets may correspond to Hα surges, and indeed, observations confirm the coexistence of x-ray jets and Hα surges (*11*, *12*, *13*).

Because magnetic reconnection is a universal physical process that can occur on any spatial or temporal scale (*14*), it should occur in the chromosphere and the photosphere (*15*), as well as the corona. The extreme-ultraviolet jets (*16*) and Hα surges (*2*, *17*) may be produced in this way in configurations like those shown in Fig. 3. If a tiny emerging bipole appears, or if a tiny bipole collides with an ambient field of opposite polarity, we can expect magnetic reconnection to occur at a relatively small spatial scale in a manner similar to that of an anemone jet observed in Yohkoh soft x-rays (*18*) (Fig. 3C). Hence, the observed morphology of the Ca jets, such as the inverted Y-shape or anemone jet, suggests that they are revealing magnetic reconnection. [Compare the expected magnetic-field configuration and jet based on a reconnection model (Fig. 3E) with that of actual observation (Fig. 3F.] Because the plasma density is high in the low chromosphere and photosphere, the Alfvèn speed is low, on the order of 10 km s^{-1}, which is comparable to the speed of sound in those layers. The plasma temperature increases by only a factor of ≤2 because the ratio of gas pressure to magnetic pressure (β) is near unity and the radiative cooling time is short in the low chromosphere and photosphere. This is consistent with the observations of Ca jets.

Figure 4 shows that there are multiple jet events in the same location and that the apparent (projected) velocity of Ca jets is typically 10 to 20 km s^{-1}, which is the same range of values predicted for cool jets in the low chromosphere and photosphere (*18*). That is, the Alfvèn velocity in the low chromosphere is estimated to be $V_A \approx 10$ km s^{-1} $(B/100\ \mathrm{G})$ $(n/10^{15}\ \mathrm{cm^{-3}})^{-1/2}$, if we assume that the density in the low chromosphere is 10^{15} cm^{-3} and the magnetic field strength is 100 G. The relation between the Ca jets and the surrounding magnetic-field distribution also shows that some of Ca jets occur in mixed-polarity regions, or near the boundary between opposite polarities (fig. S1), supporting the magnetic reconnection mechanism.

The total energy involved in a Ca jet and the footpoint is not easy to determine because the density of Ca jets is difficult to estimate. If we assume that the density at the footpoint of Ca jets is 10^{15} cm^{-3} and its temperature is 5000 K, the total thermal energy content is $E_{thermal} \approx (3/2)nkTd^3 \approx 3 \times 10^{25}$ erg $(n/10^{15}\ \mathrm{cm^{-3}})$ $(T/5000\ \mathrm{K})$ $(d/300\ \mathrm{km})^3$, where d is the typical size of the footpoint (bright point), about 0.4 arc sec ≈ 300 km from our observations. The total magnetic energy stored in the same volume is $E_{mag} \approx (B^2/8\pi)d^3 \approx 10^{25}$ erg $(B/100\ \mathrm{G})^2(d/300\ \mathrm{km})^3$. Hence, the total stored energy at the footpoint is less than a microflare value (10^{26} erg), and smaller fluctuations occurring at the footpoint would correspond to a smaller release of energy, 10^{23} to 10^{24} erg, lying in the nanoflare regime.

The velocity of the Ca jets is typically 10 to 20 km s^{-1}. But if the jet undergoes ballistic motion, i.e., if the jet is decelerated by gravity, the maximum height of the jet would be $H_{jet,max} \approx V_{jet}^2/(2g) \approx 200$ km $(V_{jet}/10\ \mathrm{km\ s^{-1}})^2$, which is much smaller than the actual height of the jet, 2000 to 5000 km. Hence, the jet cannot reach the observed height if it undergoes ballistic motion. So, how can we solve this puzzle? There are two possibilities. One is that the jet is continuously

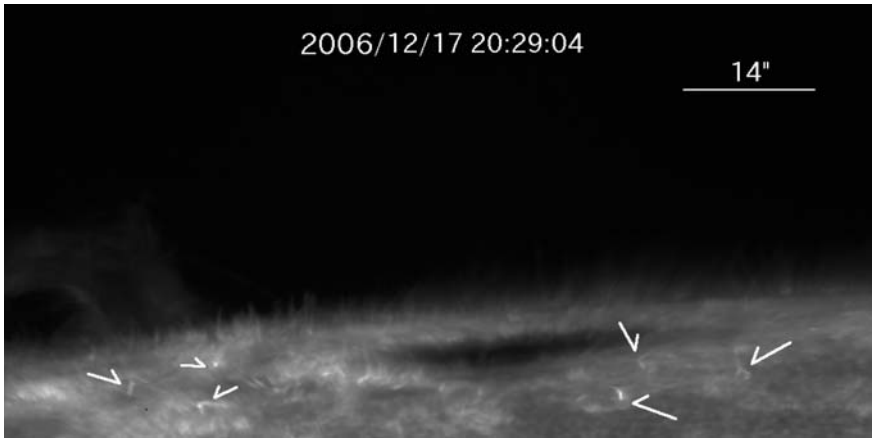


Fig. 1. Ca II H broadband filter image taken with the SOT aboard Hinode at 20:29:04 UT on 17 December 2006. For performance and examples of Ca II H observations with Hinode/SOT, see, e.g., (*29*).

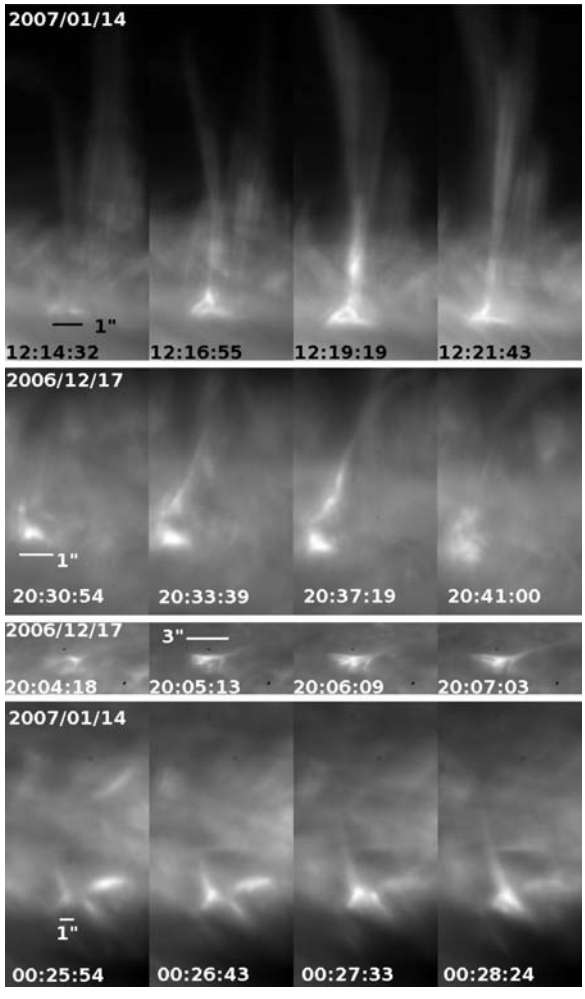


Fig. 2. Time evolution of typical Ca jets observed in Ca II H broadband filter of Hinode/SOT. Times are shown in UT.

accelerated beyond the height of 200 km. If the jet is situated along the current sheet, this would be possible. However, it is unlikely that jets with a length longer than ~1 arc sec (~700 km), such as those shown in Fig. 2, are along a current sheet longer than 1 arc sec because the bipoles at the footpoint of the jets are usually smaller than 1 arc sec (see also an example of a magnetogram in fig. S1) and because the mixed polarities at the footpoint of the jet are often formed as a result of an emerging bipole. The other possibility is that the main component of the observed Ca jet is not the plasma ejected from the footpoint but the plasma ejected from the upper chromosphere as a result of the propagation of slow-mode magneto-acoustic shocks or nonlinear Alfvèn waves (or fast-mode shocks) (Fig. 3C). Shibata *et al.* (*19*) showed that when the explosion occurs below the middle chromosphere, even if it is small so that the plasma ejected from the explosion cannot reach a higher altitude, a slow-mode shock is formed ahead of the ejecta and grows substantially. The shock accelerates the plasma in the upper chromosphere, which can reach a height of 3000 to 10000 km. Spicules and small surges ejected from Ellerman bombs may be generated in this way (*20*). The nonlinear Alfvèn wave can play a similar role to accelerate the jet from the upper chromosphere (*21*–*23*).

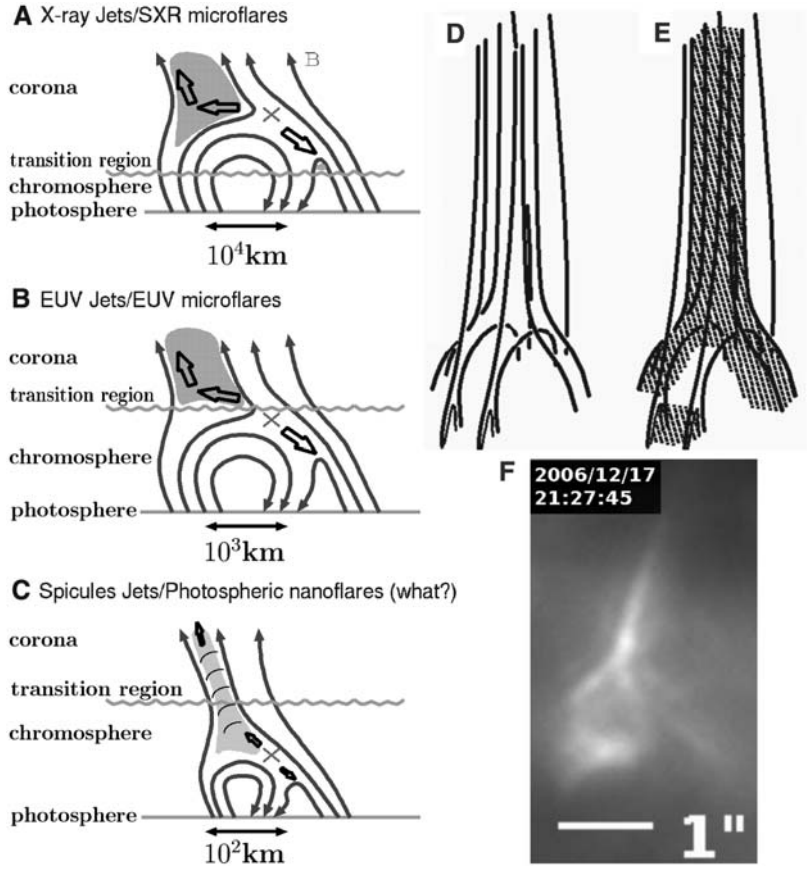


Fig. 3. (**A** to **C**) Schematic illustration of magnetic reconnections that occur at various altitudes [adapted from (*18*)]. (**D** and **E**) A model of Ca jets. (D) shows the three-dimensional magnetic-field configuration, and the hatched area in (E) shows the heated plasmas in the jet and bright point. (**F**) A typical example of an observed Ca jet.

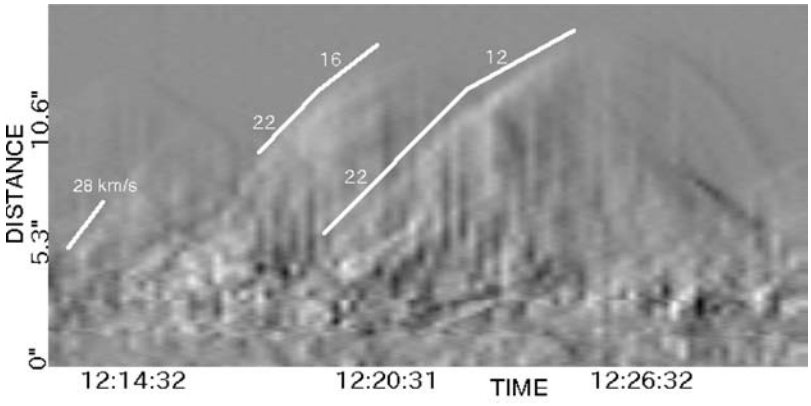


Fig. 4. Time-distance diagram of Ca intensity (running differences) along Ca jets, observed with the Ca II H broadband filter of Hinode/SOT. The numbers above the curves denote velocity (in km s^{-1}).

Parker (*4*) proposed that nanoflares (reconnection) heat the corona in quiet and active regions, and that nanoflares generate Alfvèn waves, which eventually accelerate high-speed solar winds in coronal holes (*24*, *25*). MHD simulations (*26*, *27*) showed that about 10% of energy released in the reconnection is carried away by Alfvèn waves and a comparable amount by slow-mode magnetoacoustic shocks in a typical geometry. Our finding of numerous Ca jets in the low chromosphere with indirect evidence of reconnection is consistent with these ideas.

Are these Ca jets enough to heat the corona in these regions? The total number of Ca jets in the field of view of Fig. 1 during a 1-hour period (20:00 to 21:00 UT on 17 December 2007) was 59. The total area of this field of view is roughly 50,000 km^2, so that the Ca jets occur at a rate of 7×10^{-22} s^{-1} cm^{-2}. Combining this value with the typical energy of a Ca jet of ~10^{25} erg, the average energy flux carried by these Ca jets is 7×10^3 erg cm^{-2} s^{-1}. This is much smaller than the necessary energy flux to heat the active-region corona, ~10^6 erg cm^{-2} s^{-1}. However, by definition, a Ca jet has a clear bright footpoint with an anemone shape, and a movie of the Ca II H images (movie S1, corresponding to Fig. 1) shows that almost all parts of the active-region chromosphere are covered by numerous tiny jets or jetlike features, whose footpoints are not well resolved even with Hinode. These jets seem to be continuously related to spicules, fibrils, and micro-jets above the penumbrae of sunspots (*5*), and it is likely that Ca jets are the largest events in a spectrum that has many more, smaller events. Hence, these and other Hinode data (*5*, *28*) suggest that magnetic reconnection occurs throughout the solar atmosphere and may play an important role in heating it, as conjectured by Parker (*4*).

References and Notes

1. M. Aschwanden, *Physics of the Solar Corona. An Introduction* (Springer-Verlag, Berlin, 2004).
2. D. M. Rust, in *Proceedings of the International Astronomical Union (IAU) Symposium no. 35*, Budapest, Hungary, 4 to 8 September 1967, K. O. Kiepenheuer, Ed. (Reidel, Dordrecht, Netherlands, 1968), pp. 77–84.
3. T. Kosugi *et al.*, *Solar Phys.* **243**, 3 (2007).
4. E. N. Parker, *Astrophys. J.* **330**, 474 (1988).
5. Y. Katsukawa *et al.*, *Science* **318**, 1594 (2007).
6. K. Shibata *et al.*, *Astrophys. J.* **431**, L51 (1994).
7. T. Yokoyama, K. Shibata, *Nature* **375**, 42 (1995).
8. T. Yokoyama, K. Shibata, *Publ. Astron. Soc. Jpn.* **48**, 353 (1996).
9. M. Shimojo *et al.*, *Publ. Astron. Soc. Jpn.* **48**, 123 (1996).

10. M. Shimojo, K. Shibata, K. L. Harvey, *Solar Phys.* **178**, 379 (1998).
11. K. Shibata *et al.*, *Publ. Astron. Soc. Jpn.* **44**, L173 (1992).
12. R. C. Canfield *et al.*, *Astrophys. J.* **464**, 1016 (1996).
13. B. Schmieder *et al.*, *Solar Phys.* **156**, 245 (1995).
14. E. R. Priest, T. G. Forbes, *Magnetic Reconnection* (Cambridge Univ. Press, Cambridge, 2000).
15. Y. E. Litvinenko, *Astrophys. J.* **515**, 435 (1999).
16. J. Chae *et al.*, *Astrophys. J.* **513**, L75 (1999).
17. H. Kurokawa, G. Kawai, in *Proceedings of the International Astronomical Union (IAU) Colloquium no. 141*, Beijing, China, 6 to 12 September 1992, H. Zirin, G. Ai, H. Wang, Eds. (Astronomical Society of the Pacific, San Francisco, 1993), vol. 46, pp. 507–510.
18. K. Shibata, in *The Corona and Solar Wind Near Minimum Activity*, Proc. 5th SOHO Workshop, Oslo, Norway, 17 to 20 June 1997 [European Space Agency (ESA) SP-404, 1997], pp. 103–112.
19. K. Shibata, T. Nishikawa, R. Kitai, Y. Suematsu, *Solar Phys.* **77**, 121 (1982).
20. Y. Suematsu, K. Shibata, T. Nishikawa, R. Kitai, *Solar Phys.* **75**, 99 (1982).
21. J. V. Hollweg, B. Jackson, D. Galloway, *Solar Phys.* **75**, 35 (1982).
22. T. Kudoh, K. Shibata, *Astrophys. J.* **514**, 493 (1999).
23. T. Saito, T. Kudoh, K. Shibata, *Astrophys. J.* **554**, 1151 (2001).
24. T. K. Suzuki, S. Inutsuka, *Astrophys. J.* **632**, L49 (2005).
25. W. I. Axford, J. F. McKenzie, in *Solar Wind Seven*, Proceedings of the 3rd COSPAR Colloquium, Goslar, Germany, 16 to 20 September 1991, E. Marsch, R. Schwenn, Eds. (Pergamon, Oxford, 1992), pp. 1–51.
26. A. Takeuchi, K. Shibata, *Astrophys. J.* **546**, L73 (2001).
27. T. Yokoyama, in *Solar Jets and Coronal Plumes*, Proceedings of an International Meeting, Guadeloupe, France, 23 to 26 February 1998, T.-D. Guyenne, Ed. (European Space Agency, 1998, ESA SP-421), p. 215.
28. J. W. Cirtain *et al.*, *Science* **318**, 1580 (2007).
29. T. J. Okamoto *et al.*, *Science* **318**, 1577 (2007).
30. We thank E. N. Parker, T. Magara, and S. Morita for useful comments. Hinode is a Japanese mission developed and launched by the Institute of Space and Astronautical Science/Japan Aerospace Exploration Agency, with the National Astronomical Observatory of Japan as a domestic partner and NASA and Science and Technology Facilities Council (UK) as international partners. It is operated by these agencies in cooperation with the European Space Agency and Norwegian Space Centre (Norway). This work was supported in part by the Grant-in-Aid for the 21st Century Centers of Excellence "Center for Diversity and Universality in Physics" from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and in part by the Grant-in-Aid for Creative Scientific Research "The Basic Study of Space Weather Prediction" (head investigator: K. Shibata) from the Ministry of Education, Culture, Science, Sports, and Technology of Japan.

Supporting Online Material
www.sciencemag.org/cgi/content/full/318/5856/1591/DC1
Fig. S1
Movie S1

19 June 2007; accepted 6 November 2007
10.1126/science.1146708

REPORT

Small-Scale Jetlike Features in Penumbral Chromospheres

Y. Katsukawa,[1]* T. E. Berger,[2] K. Ichimoto,[1] B. W. Lites,[3] S. Nagata,[4] T. Shimizu,[5] R. A. Shine,[2] Y. Suematsu,[1] T. D. Tarbell,[2] A. M. Title,[2] S. Tsuneta[1]

We observed fine-scale jetlike features, referred to as penumbral microjets, in chromospheres of sunspot penumbrae. The microjets were identified in image sequences of a sunspot taken through a Ca II H-line filter on the Solar Optical Telescope on board the Japanese solar physics satellite Hinode. The microjets' small width of 400 kilometers and short duration of less than 1 minute make them difficult to identify in existing observations. The microjets are possibly caused by magnetic reconnection in the complex magnetic configuration in penumbrae and have the potential to heat the corona above a sunspot.

Sunspots provide a unique opportunity for understanding plasma dynamics in strong magnetic fields. Umbral flashes, which are repetitive brightenings of the low umbral chromosphere in Ca II H and K and other spectral lines with roughly 3-min periodicity, are one of the well-known dynamic phenomena observed in the sunspot chromosphere (*1*–*3*). Penumbral fibrils seen in the Hα spectral line extend radially outward from sunspots visible in the photosphere, and an inward reversed Evershed flow is found to be associated with these dark penumbral fibrils (*4*–*6*). Running penumbral waves are observed in Hα and Ca II H and K spectral lines as a disturbance moving outward from the outer edge of an umbra and are suggested to have the same origin as the umbral flashes (*7*, *8*).

The Solar Optical Telescope (SOT) (*9*) on board Hinode (*10*) allows us to observe dynamical activities in the photosphere and chromosphere with a high and stable image quality of 0.2 arc sec (or 150 km on the solar surface) that is unaffected by atmospheric distortion. This superior performance of the SOT has led to new findings of fine-scale transient activities in the sunspot chromosphere.

We obtained image sequences of a sunspot in active region 10923 by the SOT in November 2006 through an interference filter centered at 3968 Å with a bandwidth of about 2 Å, which is sensitive to the Ca II H line. This spectral line is used to observe the lower chromosphere, where the temperature is below 10^4 K. The northeast side of the sunspot on 10 November 2006, when the spot was located at heliographic coordinates S06E60, includes a number of dark filaments extending radially in the penumbra (Fig. 1A). Umbral flashes are clearly identified in the umbra in the Ca II H movie (movie S1), and running penumbral waves are seen in the inner penumbra. The above structures and activities are all known from previous observations.

The new findings with the SOT reveal small-scale thin jetlike features in the penumbral chromosphere observed through the Ca II H filter. Because the jetlike features (hereafter referred to as penumbral microjets) are so transient, they become more prominent when viewed as a running-difference image (Fig. 1B and movie S1). The penumbral microjets appear as transient brightenings, with a brightness about 10 to 20% enhanced compared with that of the underlying penumbral structures; they become invisible several tens of seconds after their emergence. They are ubiquitous in the penumbral chromosphere, as can be seen in the Ca II H movies of sunspots (movies S1 and S2). Their length is typically between 1000 and 4000 km, and some have a length up to 10,000 km. Their width is about 400 km or less. The smallest microjet might not be resolved completely because of the spatial resolution of the SOT. The most striking property of microjets is their duration of brightening. Almost all of the penumbral microjets have durations shorter than 1 min, which is much shorter than the time scales of the known dynamical phenomena mentioned above. The small spatial and temporal scales of microjets may make it difficult to identify the phenomenon in existing ground-based observations.

When the sunspot is located far from the disk center (Fig. 1A and movie S1), the orientation of the penumbral microjets is clearly deviated from that of the penumbral filaments adjacent to the microjets in the image. The different orientation makes it easier to distinguish the penumbral microjets from the penumbral filaments in the Ca II H image. In contrast, when the spot is located close to the center of the solar disk (Fig. 1D and movie S2), as occurred on 14 November 2006, the penum-

[1]National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan. [2]Lockheed Martin Solar and Astrophysics Lab, 3251 Hanover Street, Org. ADBS, Building 252, Palo Alto, CA 94304, USA. [3]High Altitude Observatory, National Center for Atmospheric Research, Post Office Box 3000, Boulder, CO 80307, USA. [4]Kwasan and Hida Observatory, Kyoto University, Yamashina, Kyoto 607-8471, Japan. [5]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Sagamihara, Kanagawa 229-8510, Japan.

*To whom correspondence should be addressed. E-mail: yukio.katsukawa@nao.ac.jp

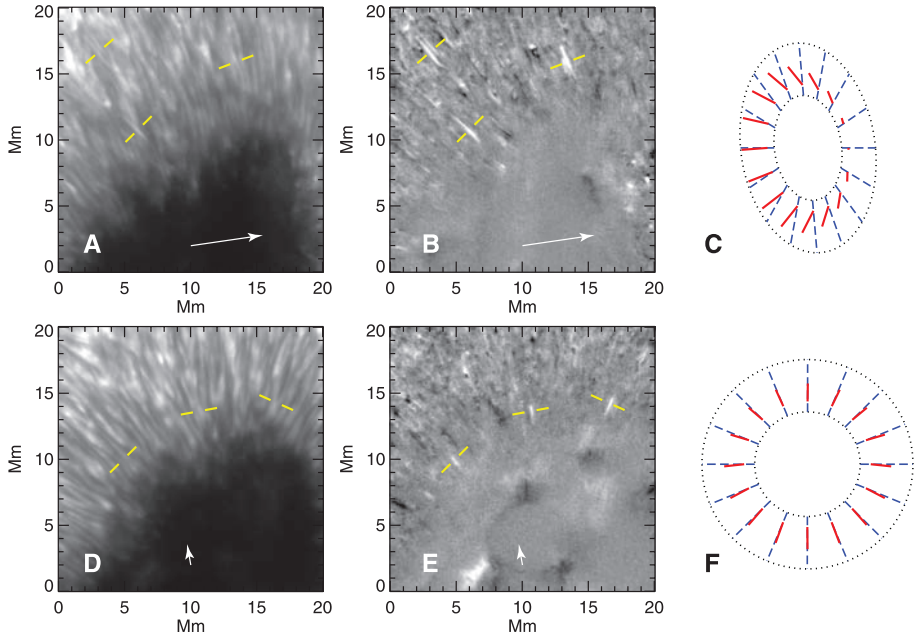


Fig. 1. (**A**) Northeast side of a sunspot taken through the Ca II H filter on SOT when the spot was located at heliographic coordinates S06E60, which was far from the center of the solar disk, and (**B**) its brightness difference from a previous frame 30 s before. (**D**) The same spot when it was located at heliographic coordinates S06E00, which was near the center of the solar disk, and (**E**) its brightness difference from a previous frame 20 s before. The yellow dashed lines indicate locations of the penumbral microjets. The arrows points to the center of the solar disk. (**C** and **F**) Schematic figures illustrating the three-dimensional configuration of the penumbral microjets when the sunspot is viewed from an oblique angle (C) and from directly above (F). The radial blue dashed lines represent penumbral filaments that are horizontal to the surface, and the red solid lines represent the penumbral microjets.

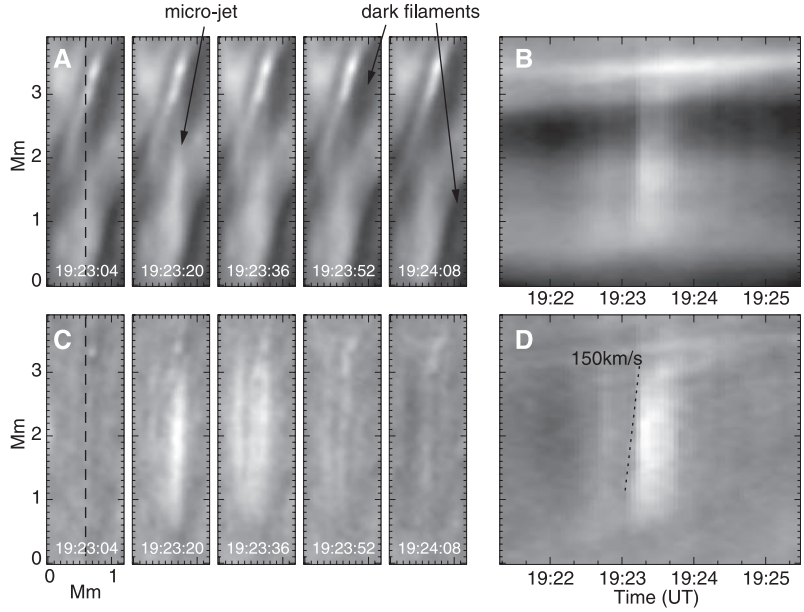


Fig. 2. (**A**) Close-up view of penumbral microjets. The images were taken 16 s apart from each other. (**B**) Space-time plot (with time increasing horizontally) with the spatial section indicated by the vertical dotted line in (A). (**C** and **D**) Same as (A) and (B) but with images processed using a temporal high-pass filter with a cutoff frequency of 3 mHz. The dotted line in (D) indicates a slope of 150 km/s to show the apparent rise velocity of the penumbral microjet.

bral microjets are difficult to identify in the Ca II H image although they are still visible in a running-difference image (Fig. 1E and movie S2). This different appearance of the penumbral microjets implies a three-dimensional configuration of the microjets and penumbral filaments. In a simple geometrical model of the penumbral microjets, when a sunspot is located far from the disk center and close to the center, the microjets are almost parallel to the penumbral filaments in the radial direction, but are more vertical to the surface while the penumbral filaments are nearly horizontal (Fig. 1, C and F). Because the microjets are not as bright as the underlying penumbral structures in the Ca II H images, it is difficult to identify them when the microjets and the penumbral filaments are aligned with each other on the projection plane. We estimated the angles of elevation of the penumbral microjets using differences of orientation angles between a penumbral filament and a penumbral microjet, taking into the account the viewing angle of the spot. The estimated elevation angles range from 20° to 60°, and most of them are between 30° and 50°.

The penumbra in the Ca II H image consists of many dark filaments, which are relatively stationary structures. There are bright structures in between two dark filaments, and penumbral microjets emanate from there (Fig. 2A). The microjets are easily identified in the high-pass–filtered images (Fig. 2C). The penumbral microjets begin to brighten suddenly and darken gradually in several tens of seconds. Evolution of the brightness of the microjet after emergence is relatively uniform along its length, but the brightening seems to start from the root of the microjet and then peaks after about 10 s (Fig. 2D). The evolution of the intensity pattern provides an apparent rise velocity faster than 100 km/s in the initial phase (Fig. 2D). The velocity is much faster than the acoustic velocity in the chromosphere, which is about 10 km/s.

Space-time plots (Fig. 3, A and B) covering a longer period illustrate the time development of locations of the penumbral microjets along a bright lane in the penumbra. There are three kinds of dynamical phenomena in the high-pass–filtered plot (Fig. 3B). Running penumbral waves are clearly seen propagating outward in the inner penumbra with a period of about 3 min. Their propagating velocity is about 18 km/s, which is comparable to or slightly faster than the acoustic velocity in the chromosphere (*7*, *8*). In the outer penumbra, many tiny features, about 200 km in size, are observed moving outward with velocities of 5 to 10 km/s. Those features might be associated with the Evershed flow observed in the photosphere (*11*, *12*). Vertical bright strips in the space-time plot (Fig. 3B) correspond to the penumbral microjets. It is clear that the penumbral microjets are gradually drifting inward to the umbra. The inward motion of a bright region can be seen in the unfiltered plot (Fig. 3A). The bright

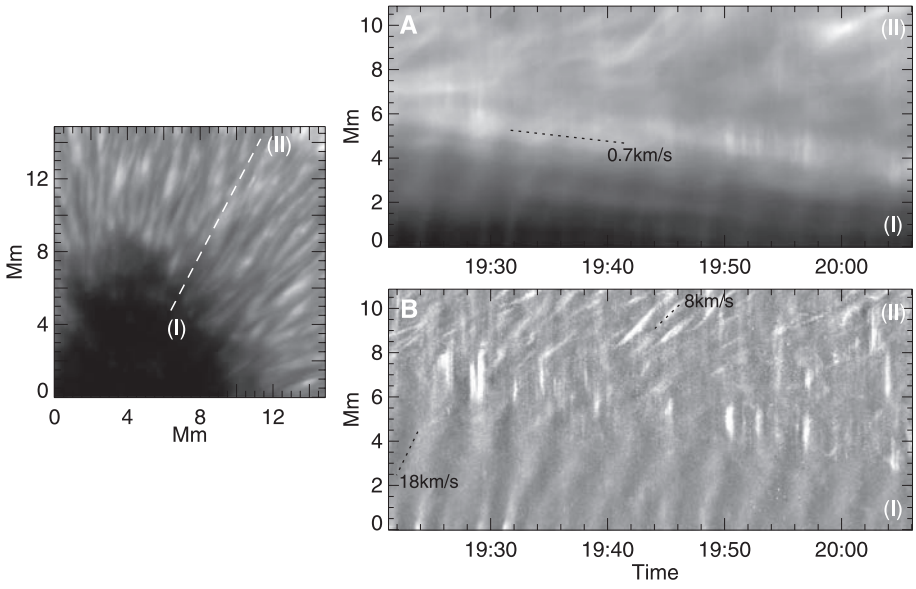


Fig. 3. (**A**) Space-time plots (with time increasing horizontally) with the spatial cut along a bright lane in the penumbra indicated by the dashed line between (I) and (II) in the left panel, which is a Ca II H image at the center of the time range. (**B**) Same as (A) but with images processed using a temporal high-pass filter with a cutoff frequency of 3 mHz. The dotted line in (A) indicates a slope of 0.7 km/s to show the inward velocity of a penumbral grain. The two dotted lines in (B) indicate slopes of 18 and 8 km/s to show velocities of running penumbral waves in the inner penumbra and outward-moving features in the outer penumbra, respectively.

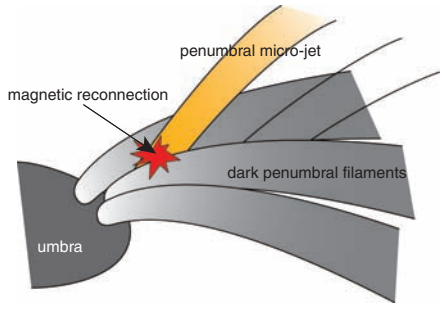


Fig. 4. Sketch showing inferred configuration of a penumbral microjet and penumbral dark filaments. There are nearly horizontal magnetic fields along the dark penumbral filaments, and the penumbral microjet has its root in between the two dark penumbral filaments, where more vertical magnetic fields exist. The magnetic configuration possibly induces magnetic reconnection between the horizontal and vertical fields and generates transient jetlike brightenings.

structure is presumably a penumbral grain, which is known to be associated with inward-migrating motions (*13*, *14*). The penumbral microjets are found to be located near the penumbral grains.

The observed properties of the penumbral microjets have similarities to jetlike activities seen in the chromosphere outside of sunspots. These activities referred to as spicules along the limb, as dynamic fibrils in active-region plages, and as mottles in the quiet Sun (*15*–*18*). Their spatial scales are 3 to 10 Mm in height and 0.2 to 1 Mm in width, which are similar to the spatial scale of the penumbral microjets. One difference is that the jetlike activities outside of sunspots have ascending and descending motion with velocities of 10 to 30 km/s during their lifetimes of 3 to 15 min. The velocities are substantially slower and the lifetimes longer than those of the penumbral microjets. This implies a different driver for each of these jetlike activities.

The magnetic configuration of sunspot penumbrae has been extensively studied using filter magnetograms (*19*, *20*) and spectro-polarimetric measurements (*21*–*23*) of spectral lines forming in the photosphere. Those studies indicate horizontal magnetic fields along dark penumbral filaments and more vertical fields in bright regions of penumbrae (*20*). If the two magnetic components with different vectors closely face each other, large electric currents are naturally generated between them, resulting in rapid dissipation of magnetic energies through magnetic reconnection (Fig. 4). The inward motion of penumbral grains may help the formation of strong current sheets at a leading edge of a penumbral filament. Magnetic reconnection is a well-accepted mechanism of flares and mass ejections in the corona (*24*) and might explain the transient nature of the penumbral microjets. The thermal energy content of each event can be estimated by $E = (3/2)nk_BTV$, where n is the number density, k_B is the Boltzmann constant, T is the temperature, and V is the volume. When $T = 10^4$ K and $V = 2000$ km × (300 km)^2, we find that $E = 2 \times 10^{16}$ J, or 2×10^{23} erg for $n = 10^{18}$ m^{-3}. This energy is comparable to the lowest energy of nanoflares observed in the corona (*25*, *26*). The penumbral microjets are potentially involved in heating the transition region and corona above the sunspots.

There are two possible explanations for the apparent high-speed lateral motion observed in the Ca II H image sequence. One is true mass motion with very high velocities exceeding the acoustic velocity. The other possibility is that the jetlike brightenings are a signature of plasma transiently heated to much higher temperatures. The microjets become visible after their temperatures are cooled down to below 10^4 K, at which point the Ca II H line is formed. In the former case, the high-speed motion might be explained by a reconnection outflow with an Alfvén velocity that exceeds 200 km/s when $B = 100$ G and $n = 10^{18}$ m^{-3}. In the latter case, the apparent motion is not the true mass motion, but is due to the evolution of a thermal conduction front. A Doppler shift measurement of the penumbral microjets is essential to determine which explanation is correct.

References and Notes

1. J. M. Beckers, P. E. Tallant, *Solar Phys.* **7**, 351 (1969).
2. A. Wittmann, *Solar Phys.* **7**, 366 (1969).
3. J. H. Thomas, L. E. Cram, A. H. Nye, *Astrophys. J.* **285**, 368 (1984).
4. P. Foukal, *Solar Phys.* **20**, 298 (1971).
5. P. Maltby, *Solar Phys.* **43**, 91 (1975).
6. C. E. Alissandrakis, D. Dialetis, P. Mein, B. Schmieder, G. Simon, *Astron. Astrophys.* **201**, 339 (1988).
7. G. Tsiropoula, C. E. Alissandrakis, P. Mein, *Astron. Astrophys.* **355**, 375 (2000).
8. L. H. M. Rouppe van der Voort *et al.*, *Astron. Astrophys.* **403**, 277 (2003).
9. K. Ichimoto *et al.*, *Proc. SPIE* **5487**, 1142 (2004).
10. T. Kosugi *et al.*, *Solar Phys.* **243**, 3 (2007).
11. R. A. Shine *et al.*, *Astrophys. J.* **430**, 413 (1994).
12. T. R. Rimmele, *Astron. Astrophys.* **290**, 972 (1994).
13. R. Muller, *Solar Phys.* **29**, 55 (1973).
14. M. Sobotka, P. N. Brandt, G. W. Simon, *Astron. Astrophys.* **348**, 621 (1999).
15. Y. Suematsu, H. Wang, H. Zirin, *Astrophys. J.* **450**, 411 (1995).
16. K. Tziotziou, G. Tsiropoula, P. Sütterlin, *Astron. Astrophys.* **444**, 265 (2005).
17. R. J. Rutten, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, R. F. Stein, H. Uitenbroek, Eds. (Astronomical Society of the Pacific, San Francisco, 2006), vol. 354, pp. 276–283.
18. B. De Pontieu, V. H. Hansteen, L. Rouppe van der Voort, M. van Noort, M. Carlsson, *Astrophys. J.* **655**, 624 (2007).
19. A. M. Title *et al.*, *Astrophys. J.* **403**, 780 (1993).
20. K. Langhans, G. B. Scharmer, D. Kiselman, M. G. Löfdahl, T. E. Berger, *Astron. Astrophys.* **436**, 1087 (2005).
21. B. W. Lites, D. F. Elmore, P. Seagraves, A. P. Skumanich, *Astrophys. J.* **418**, 928 (1993).
22. V. Martínez Pillet, *Astron. Astrophys.* **361**, 734 (2000).
23. L. R. Bellot Rubio, H. Balthasar, M. Collados, *Astron. Astrophys.* **427**, 319 (2004).
24. K. Shibata, *Adv. Space Res.* **17**, 9 (1996).
25. M. J. Aschwanden *et al.*, *Astrophys. J.* **535**, 1047 (2000).
26. Y. Katsukawa, *Publ. Astron. Soc. Jpn.* **55**, 1025 (2003).

27. We thank K. Shibata and J. Jurcak for suggestions and helpful comments on this study. We also thank the late Professor T. Kosugi for his great contribution to the success of the Hinode mission. Hinode is a Japanese mission developed and launched by the Institute of Space and Astronautical Science/Japan Aerospace Exploration Agency,

with the National Astronomical Observatory of Japan (NAOJ) as domestic partner and NASA and Science and Technology Facilities Council (UK) as international partners. It is operated by these agencies in cooperation with the European Space Agency and the Norwegian Space Centre. This work was carried out at the NAOJ Hinode Science Center, which is supported by the Grant-in-Aid for Creative Scientific Research, "The Basic Study of Space Weather Prediction" (head investigator: K. Shibata), from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; a generous donation from Sun Microsystems Inc.; and NAOJ internal funding.

Supporting Online Material
www.sciencemag.org/cgi/content/full/318/5856/1594/DC1
Movies S1 and S2

4 June 2007; accepted 1 November 2007
10.1126/science.1146046

REPORT

Twisting Motions of Sunspot Penumbral Filaments

K. Ichimoto,[1]* Y. Suematsu,[1] S. Tsuneta,[1] Y. Katsukawa,[1] T. Shimizu,[2] R. A. Shine,[3] T. D. Tarbell,[3] A. M. Title,[3] B. W. Lites,[4] M. Kubo,[4] S. Nagata[5]

The penumbra of a sunspot is composed of numerous thin, radially extended, bright and dark filaments carrying outward gas flows (the Evershed flow). Using high-resolution images obtained by the Solar Optical Telescope aboard the solar physics satellite Hinode, we discovered a number of penumbral bright filaments revealing twisting motions about their axes. These twisting motions are observed only in penumbrae located in the direction perpendicular to the symmetry line connecting the sunspot center and the solar disk center, and the direction of the twist (that is, lateral motions of intensity fluctuation across filaments) is always from limb side to disk-center side. Thus, the twisting feature is not an actual twist or turn of filaments but a manifestation of dynamics of penumbral filaments with three-dimensional radiative transfer effects.

A sunspot's penumbra, an annular region surrounding the sunspot's dark core (umbra), provides us with an opportunity to study the dynamical processes of strongly magnetized plasma in detail. Sunspot penumbrae exhibit fine-scale structures consisting of a number of radial bright and dark filaments having widths of 0.2″ (150 km on the Sun) or less (*1*–*3*). A nearly horizontal outward gas flow, known as the Evershed flow (*4*), takes place in the photospheric layer of the penumbra and is closely related to the fine structures of the penumbra. The flow is concentrated in dark filaments, at least in the outer penumbra, where the magnetic field is more horizontal than in other parts of the penumbra (*5*–*13*). Penumbral grains, inwardly migrating bright structures preferentially found in the inner penumbra (*14*), are identified as the inner footpoints of Evershed flows, where a hot upflow occurs (*15*, *16*). Penumbral filaments consist of a dark core flanked by lateral brightening (*17*–*20*). This structure behaves as a single entity and is thought to be a fundamental structure of penumbrae, providing an important clue for understanding the nature of the penumbra.

A number of models have been proposed to account for the filamentary structure of penumbrae. Three scenarios recently at the center of the debate over the nature of penumbral fine structure are (i) the embedded flux tube model (*21*, *22*), in which rising magnetic flux tubes embedded in more-vertical background magnetic fields in the penumbra carry the Evershed flow; (ii) the gappy penumbral model (*23*, *24*), in which the penumbral bright filaments are regarded as protrusions of nonmagnetized convecting hot gas into the background oblique magnetic fields of the penumbra; and (iii) a roll-like magnetoconvection model with weaker (but still finite) magnetic fields in the rising and expanding plumes (*25*), as demonstrated in realistic magnetohydrodynamic simulations (*26*, *27*). There is still no consensus on the physical nature of penumbrae.

The Solar Optical Telescope (SOT) aboard Hinode, the third Japanese solar physics satellite (*28*, *29*), obtains diffraction-limited images with a 50-cm aperture telescope from a Sun-synchronous orbit. The SOT provides us with an opportunity for studying small-scale dynamics on the Sun with high image stability. In this paper, we present a new feature found in sunspot penumbra: twisting motions of penumbral filaments.

The data were obtained on 12 November 2006 (Fig. 1). The sunspot (NOAA10923), located at 28°E, 7°S (at a heliocentric angle of ~30°), had a well-developed penumbra. We used blue continuum images at 4500 Å taken by the SOT Broadband Filter Imager (BFI). The spatial resolution of the BFI is about 0.2 arcsec with a pixel scale of 0.054 arcsec per pixel. The time cadence of the data set is 20 s. Apparent motions of fine structures in the penumbra were investigated with a movie made from the time series of the images ~95 min in duration (movie S1). The spectropolarimeter (SP) of the SOT took full Stokes profiles (linear and circular polarizations) of the pair of FeI lines at 6301.5 and 6302.5 Å with a photometric accuracy of 0.1% and a pixel resolution of 0.16 arcsec per pixel. A map of the entire sunspot made with the SP provides the spatial distribution of the line-of-sight (LOS) velocity.

One obvious feature (movie S1) is the intensity fluctuations in penumbral filaments, showing evidence of outward apparent motions along the filaments. The outward motions are also visible even in inwardly migrating bright filaments in the inner penumbra. This behavior reinforces the finding that the penumbral bright grains are the footpoints of the Evershed flows (*15*), and that the flows start from the innermost edge of the bright filaments. Another noticeable feature is that some bright filaments reveal internal intensity fluctuations moving laterally across the filaments, giving rise to the strong impression of twisting motions of individual filaments. Narrow dark lanes that are inclined toward or nearly parallel to the axis of the bright

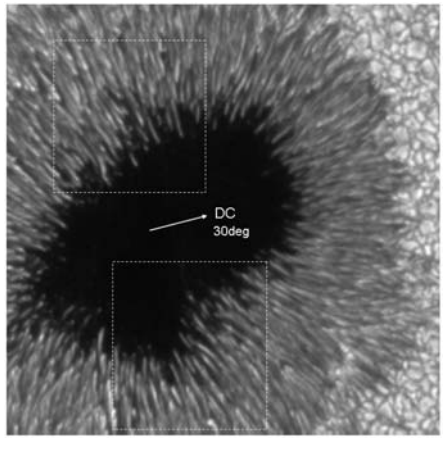


Fig. 1. A blue continuum image of the sunspot on 12 November 2006 taken by the SOT BFI aboard Hinode. Top and right are toward solar north and west, respectively. The arrow at the center of the sunspot indicates the direction toward the solar disk center. Two boxes enclosed by dashed lines show the subfields, which are presented in movie S1.

[1]National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Toyko 181-8588, Japan. [2]Japan Aerospace Exploration Agency, Institute of Space and Astronautical Science, 3-1-1, Yoshinodai, Sagamihara, Kanagawa 229-8510, Japan. [3]Lockheed Martin Advanced Technology Center, 3251 Hanover Street, Palo Alto, CA 94304, USA. [4]High Altitude Observatory, National Center for Atmospheric Research, Post Office Box 3000, Boulder, CO 80307-3000, USA. [5]Kwasan and Hida Observatories, Kyoto University, Yamashina, Kyoto 607-8471, Japan.

*To whom correspondence should be addressed. E-mail: ichimoto@solar.mtk.nao.ac.jp

filaments move from one side of the filaments to the other. This behavior was also noted in (*16*). Also, filaments belonging to the same section of the sunspot twist in the same direction, but those in northern and southern sections twist in opposite directions.

To better demonstrate the lateral movement of intensity fluctuations across the filaments, space-time plots were created along lines crossing the filaments in the inner penumbra (Fig. 2). A number of bright filaments revealed twisted ropelike structures. The typical speed of the apparent lateral motion of the intensity across the filament was 1 to 2 km/s. The twisting motion, if present in the filament, was persistent throughout our observed time span of about 95 min, with a period for the brightness fluctuations of 3 to 5 min. The common direction of the twist among the filaments in the same section of the sunspot, and the opposite directions of the twist between different sections, were also confirmed in the space-time plot.

To examine the physical nature of this apparent twist of the filaments, the LOS velocities across the filaments were investigated using the SP data (Fig. 3). The LOS velocity is calculated from the wavelength position of the center of gravity of the FeI 6301.5 Å line. It was found (Fig. 3) that there is a relationship between the intensity and the blue shift, but the position of peaks of the intensity and the blue shift show a slight offset with each other; in the northern section (Fig. 3A), the bright filaments (the intensity peak in the upper profile) are located in a positive gradient of the Doppler shift; namely, the left side of the filament is relatively blue-shifted with respect to the right side. The opposite is true for the filaments in the southern section. These signatures in Doppler shifts are consistent in sign with the interpretation of the apparent twist of the filaments as an actual twisting motion or turn of individual filaments about their axes.

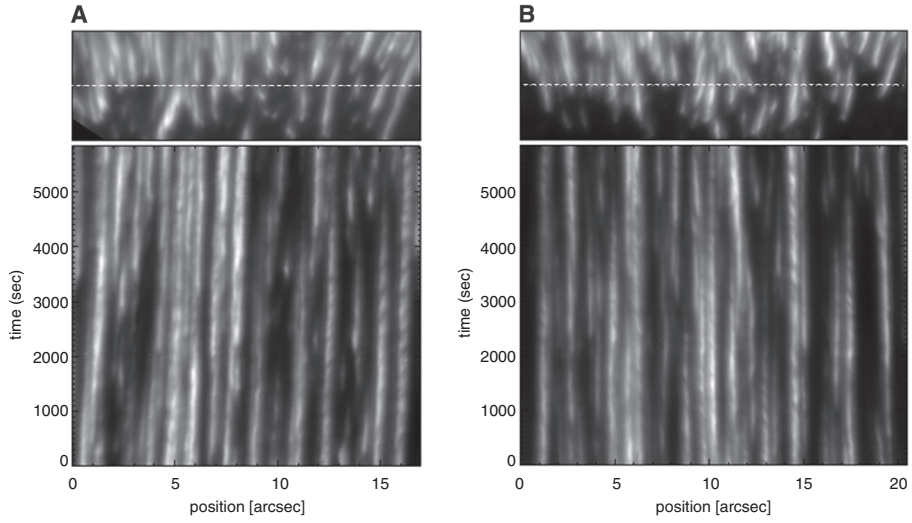


Fig. 2. Space-time plots along a line across the penumbral filaments. The top panels are snapshots of penumbral images, with a dashed line indicating the slice for the space-time plot in the bottom. (**A**) shows the northern section and (**B**) shows the southern section of the sunspot (Fig. 1).

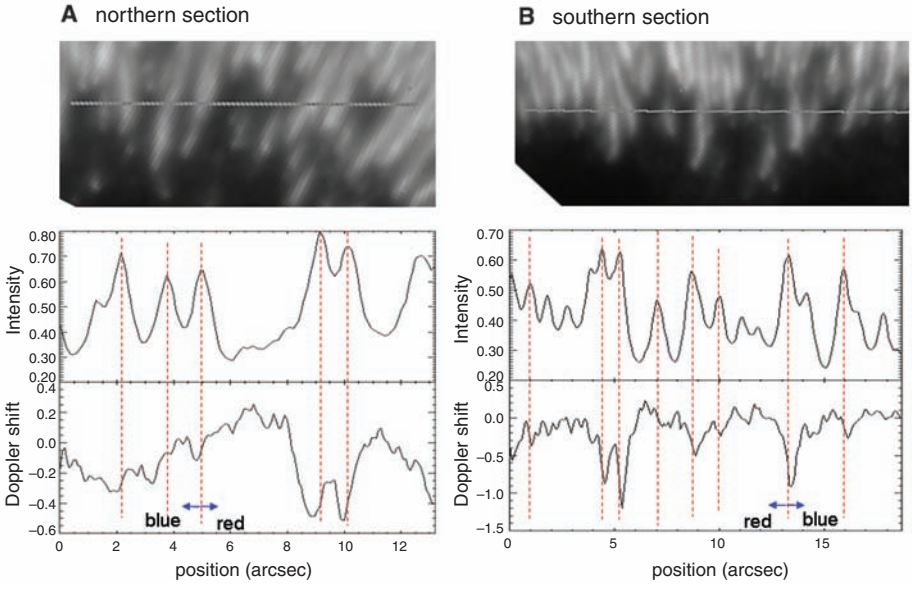


Fig. 3. Profiles of continuum intensity (middle) and Doppler shift (bottom) of the FeI 6301.5 Å line across the penumbral filament. The top panels show the line for the plots on the penumbral images. Red shift is positive and blue shift is negative in the plot of the Doppler shift. (**A**) and (**B**) show the northern and southern sections of the sunspot, respectively.

We note, however, that the apparent twisting motions of penumbral filaments are observed only in restricted sections of the sunspot; they are found in the penumbra, which is located in the direction perpendicular to the symmetry line connecting the sunspot center and the solar disk center, and are hardly seen in the limb-side and disc-center–side penumbra. This means that the direction of the twist of the filaments (or the lateral motions of intensity fluctuation in the filaments) is from the limb side toward the disk-center side for both the northern and southern sections of the sunspot. Therefore, a view angle effect may determine the sign of the apparent twisting motions of the filaments. To further examine this behavior in terms of view angle, we examined the data for this sunspot on other days (mostly with BFI images taken in G band 4300 Å and with reduced time resolution). On 10 and 11 November, when the sunspot was located in the eastern hemisphere, and on 16 November, when the sunspot was located in the western hemisphere (30°W, 7°S), we observed that the apparent twist was always directed from the limb side toward the disk center in both sections of penumbrae (Fig. 4). Furthermore, we investigated another sunspot (NOAA10933), observed on 6 to 11 January 2007, and confirmed that penumbral filaments again showed the apparent twists in directions consistent with the rule described above. Thus, we conclude that the apparent twists observed in penumbral filaments are neither an actual twist nor a helical motion of individual filaments but rather are a manifestation of their dynamical nature, so that their appearance depends on the viewing angle.

As a natural consequence of our findings, penumbral filaments showing the apparent twisting motions are not associated with the stationary dark cores. The fact that the apparent twists of the filaments are observed only in penumbral filaments located in the direction perpendicular to the symmetry line (the line connecting the sunspot center and the solar disk center) is consistent with the fact that the dark cores of penumbral filaments are most visible

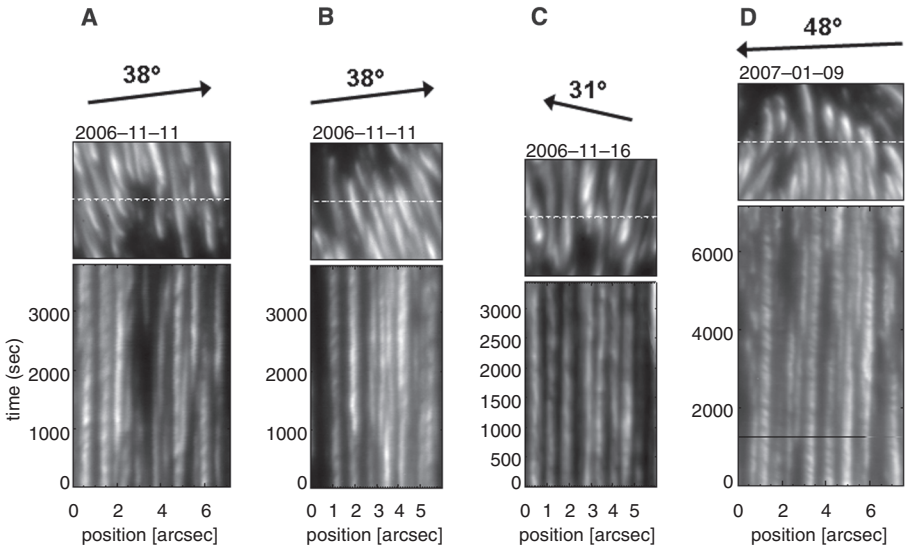


Fig. 4. (**A** to **D**) Examples of space-time slices across the inner penumbral filaments observed at different locations on the Sun. At the top of each panel, the direction and distance to the solar disk center are shown by an arrow. Dates are given at the top of each panel. The upper panels show images of penumbral filaments; the dashed line indicates the space-time slice.

where the penumbral filamentary structure is oriented along the symmetry line (*18*–*20*).

The mechanism that produces the apparent twisting motions of the penumbral filaments is not clear at this point. There may be interactions of a rising magnetic flux tube carrying a strong Evershed flow with the ambient oblique magnetic fields. Dark features on the bright penumbral filaments could be interpreted as dark structures along the ambient, more vertical magnetic fields overlying the nearly horizontal bright filaments. It is not easy, however, to imagine that these dark features move from limb side to disk-center side across the filaments. Also, this picture may not explain the observed feature of the Doppler shift associated with the bright filaments.

According to recent numerical simulations of magnetoconvection (*26*, *27*), the dark core of penumbral filaments can be explained by a locally elevated $\tau = 1$(continuum optical depth unity) surface over the constant temperature surface due to a hot and dense plume formed by an upflow of the overturning convection. When this structure is viewed from the side, the dark lane, or the upper part of the upflow plume, will be projected to the limb side and the peak of brightness will be at the disk-center–side flank of the plume. Such a geometry may explain the limbward displacement of the peak of the blue-shifted velocity with respect to the peak of the intensity (Fig. 3). The convective upflow then turns to a nearly horizontal outflow (Evershed flow), and along with the flow (or time), the dense plume probably becomes less tall and cooler because of gravity and radiative cooling. The consequent deformations of the $\tau = 1$ surface and the constant-temperature surface might cause the dark features to apparently move toward the disc center. This is a hypothetical scenario that may explain the twisting motion of the bright filaments. The recurrent intensity fluctuations in the filaments with periods of 3 to 5 min may suggest an interaction of the convective motion or the rising flux tube with the umbral oscillation or running penumbral waves.

References and Notes

1. S. Solanki, *Astron. Astrophys. Rev.* **11**, 153 (2003).
2. J. H. Thomas, N. O. Weiss, *Annu. Rev. Astron. Astrophys.* **42**, 517 (2004).
3. N. O. Weiss, *Astron. Soc. Pac. Conf. Ser.* **354**, 213 (2006).
4. J. Evershed, *Mon. Not. R. Astron. Soc.* **69**, 454 (1909).
5. A. Title *et al.*, *Astrophys. J.* **403**, 780 (1993).
6. W. Schmidt, A. Hofmann, H. Balthasar, T. Tarbell, Z. Frank, *Astron. Astrophys.* **264**, L27 (1992).
7. B. W. Lites, D. F. Elmore, P. Seagraves, A. P. Skumanich, *Astrophys. J.* **418**, 928 (1993).
8. K. Langhans, G. B. Scharmer, D. Kiselman, M. G. Lofdahl, T. E. Berger, *Astron. Astrophys.* **436**, 1087 (2005).
9. L. R. Bellot Rubio, H. Balthasar, M. Collados, *Astron. Astrophys.* **427**, 319 (2004).
10. J. M. Beckers, E. H. Schroter, *Sol. Phys.* **10**, 384 (1969).
11. R. A. Shine, A. Title, T. Tarbell, K. Smith, Z. Frank, *Astrophys. J.* **430**, 413 (1994).
12. R. Schlichenmaier, L. R. Bellot Rubio, A. Tritschler, *Astron. Nachr.* **326**, 301 (2005).
13. L. R. Bellot Rubio, R. Schlichenmaier, A. Tritschler, *Astron. Astrophys.* **453**, 1117 (2006).
14. R. Muller, *Sol. Phys.* **48**, 101 (1976).
15. T. Rimmele, J. Marino, *Astrophys. J.* **646**, 593 (2006).
16. K. Ichimoto *et al.*, *Publ. Astron. Soc. Jpn.* **59**, 593 (2007).
17. G. B. Scharmer, B. V. Gudiksen, D. Kiselman, M. G. Lofdahl, L. H. M. Rouppe van der Voort, *Nature* **420**, 151 (2002).
18. P. Sutterlin, L. R. Bellot Rubio, R. Schlichenmaier, *Astron. Astrophys.* **424**, 1049 (2004).
19. L. R. Bellot Rubio, K. Langhans, R. Schlichenmaier, *Astron. Astrophys.* **443**, L7 (2005).
20. K. Langhans, G. B. Scharmer, D. Kiselman, M. G. Lofdahl, *Astron. Astrophys.* **464**, 763 (2007).
21. S. K. Solanki, C. A. P. Motavon, *Astron. Astrophys.* **275**, 283 (1993).
22. R. Schlichenmaier, K. Jahn, H. U. Schmidt, *Astron. Astrophys.* **337**, 897 (1998).
23. H. C. Spruit, G. B. Scharmer, *Astron. Astrophys.* **447**, 343 (2006).
24. G. B. Scharmer, H. C. Spruit, *Astron. Astrophys.* **460**, 605 (2006).
25. N. O. Weiss, D. P. Brownjohn, N. E. Hurburt, M. R. Proctor, *Mon. Not. R. Astron. Soc.* **245**, 434 (1990).
26. M. Schussler, A. Vogler, *Astrophys. J. Lett.* **641**, L73 (2006).
27. T. Heinemman, A. Nordlund, G. B Scharmer, H. C. Spruit, *Astrophys. J.*, preprint available at arXiv:astro-ph/0612648v2 (2007).
28. K. Ichimoto and the Solar-B Team, *J. Korean Astron. Soc.* **38**, 307 (2005).
29. T. Kosugi *et al.*, *Sol. Phys.*, **243**, 3 (2007).

30. The authors are grateful to the late professor T. Kosugi of the Japan Aerospace Exploration Agency/Institute of Space and Astronautical Science (JAXA/ISAS) and to L. Hill, R. Jayroe, and J. Owens of NASA for continuous support throughout the development of the SOT. Hinode is a Japanese mission developed and launched by JAXA/ISAS, with the National Astronomical Observatory of Japan as domestic partner and NASA and Science and Technology Facilities Council (UK) as international partners. It is operated by these agencies in cooperation with the European Space Agency and Norwegian Space Centre (Norway). The National Center for Atmospheric Research is sponsored by NSF. The referee's comments greatly helped us to improve the manuscript, especially for the paragraph of interpretations and discussions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1597/DC1
Movie S1

11 June 2007; accepted 1 November 2007
10.1126/science.1146337

Economics of Overexploitation Revisited

R. Q. Grafton,[1]* T. Kompas,[1] R. W. Hilborn[2]

About 25% of the world's fisheries are depleted (*1*) such that their current biomass (B_{CUR}) is less than the biomass that maximizes the sustained yield (B_{MSY}). We show that, if B_{CUR} were compared to the biomass that produces the largest discounted economic profits from fishing (B_{MEY}), many more stocks would be considered overexploited, i.e., $B_{MEY} > B_{MSY}$ and $B_{MEY} > B_{CUR}$.

A classic result (*2*) is that it may be optimal for a private owner to overexploit to extinction a renewable resource with low reproductive capacity, whereas B_{MEY} may be either greater or smaller than B_{MSY} depending on the discount rate, sensitivity of costs and revenues to biomass and harvest, and the marginal growth in the biomass (*3*). We show that under reasonable prices, costs, and discount rates it is not economic to exploit fisheries to extinction even with very low growth rates. Moreover, we found that B_{MEY} exceeds B_{MSY}. With use of perturbation methods (*4*), we solved for optimal stochastic harvest and biomass transition paths to B_{MEY}, specified revenues as a nonlinear function of harvest, and allowed for a "stock effect" such that costs are a nonlinear function of the biomass.

Model outputs for four different fisheries are presented (Fig. 1) where equilibrium revenue and economic profit curves (in thousands of U.S. dollars) are plotted against equilibrium fish biomass or spawning stock (in tons). We modeled population dynamics with an age-structured model but used a spawner-recruitment model in the prawn fishery (*5*). The peak of the revenue curves corresponds to B_{MSY}, whereas the maximum of the economic profit curves equals B_{MEY}. We found that $B_{MSY} > B_{CUR}$ except for yellowfin tuna, but in all cases B_{MEY} exceeded B_{MSY} at a discount rate of 5%.

The profit two curve shows the sensitivity of the results to costs [extra 30 cents (U.S.) per liter of fuel] relative to the base-case profit one curve. Higher costs, lower output prices, and smaller discount rates increase the ratio of B_{MEY} to B_{MSY}. The profit curves show that the more overexploited the fishery, the greater the marginal economic gains from stock rebuilding. Although the stock effect may be relatively small in some cases (*6*), our estimates indicate that it is large at current biomass, whereas harvesting costs rise at an increasing rate with declines in the biomass. We also find that a discount rate as large as 25% for the tuna and prawn fisheries and a discount rate as high as 10% for orange roughy still generate $B_{MEY} > B_{MSY}$. In all cases, even at very high discount rates, fishing profits become negative long before biomass is fully depleted.

Theoretically, maximizing discounted economic profits may cause stock depletion or even extinction (*2*). Our results show, however, that in practice $B_{MEY} > B_{MSY}$. This implies a win-win: Conservation promotes both larger fish stocks and higher profits. When B_{MEY} exceeds B_{CUR}, the debate is no longer whether it is economically advantageous to reduce current harvests but how fast stocks should be rebuilt.

Revisiting the economics of overexploitation provides both a target and a framework to help overcome a key cause of overharvesting: fisher opposition to lower catches associated with stock rebuilding. For fast-growing species, rebuilding to B_{MEY} may only take a short time, but for very slow-growing fish it may take decades. Transition costs to B_{MEY} may explain why fishers oppose stock rebuilding, as do stakeholder conflicts over whether to maximize short-run employment, to maximize discounted economic profits, or to achieve ecosystem conservation goals (*7*). Appropriate incentives in the form of individual or community harvesting rights (*8*) can provide security to current fishers that they will benefit from stock rebuilding. Intertemporal transfers from higher, future profits in the form of a quota rental charge (*9*) could also compensate for transition costs and employment losses in moving to B_{MEY}.

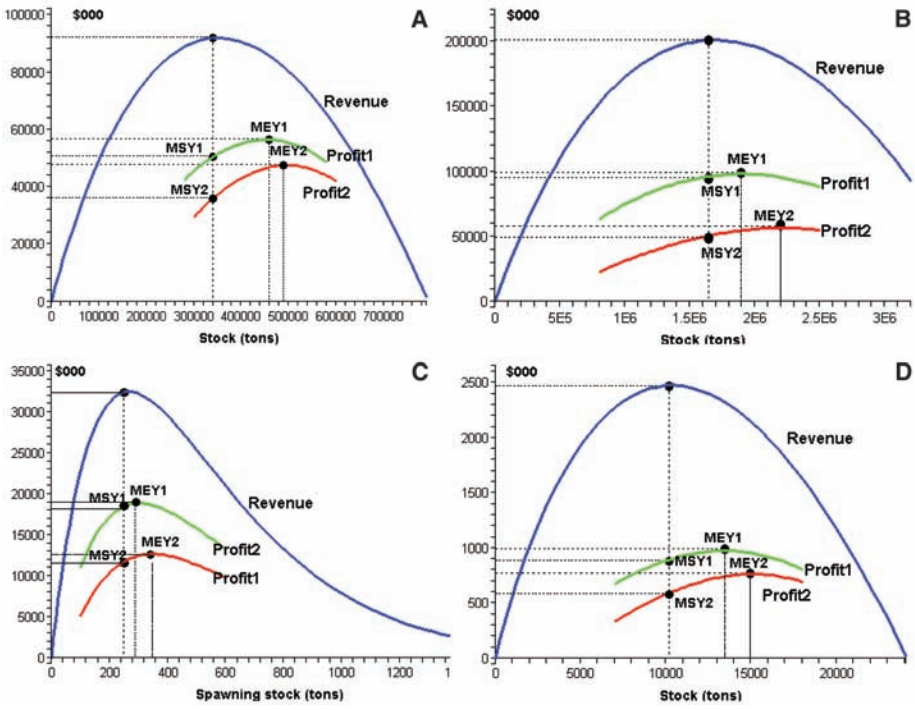


Fig. 1. (**A**) B_{MEY} and B_{MSY} of Western and Central Pacific big eye tuna. (**B**) B_{MEY} and B_{MSY} of Western and Central Pacific yellowfin tuna. (**C**) B_{MEY} and B_{MSY} of Australian northern prawn fishery. (**D**) B_{MEY} and B_{MSY} of Australian orange roughy fishery.

References and Notes

1. Food and Agriculture Organization (FAO), *State of the World Fisheries and Aquaculture 2006* (FAO, Rome, 2007).
2. C. W. Clark, *Science* **181**, 630 (1973).
3. C. W. Clark, G. R. Munro, *J. Environ. Econ. Manag.* **2**, 92 (1975).
4. K. L. Judd, *Numerical Methods for Economics* (MIT Press, Cambridge, MA, 1999).
5. T. J. Quinn II, R. B. Deriso, *Quantitative Fish Dynamics* (Oxford Univ. Press, New York, 1999).
6. R. Hannesson, *Mar. Res. Econ.* **22**, 69 (2007).
7. R. Hilborn, *Mar. Policy* **31**, 153 (2007).
8. R. Q. Grafton *et al.*, *Can. J. Fish. Aquat. Sci.* **63**, 699 (2006).
9. R. Q. Grafton, *J. Environ. Econ. Manag.* **28**, 48 (1995).

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1601/DC1
Materials and Methods
References

4 June 2007; accepted 10 October 2007
10.1126/science.1146017

[1]Crawford School, Australian National University, Ellery Crescent, Canberra, ACT 0200, Australia. [2]School of Aquatic and Fishery Sciences, Box 355020, University of Washington, 112 Boat Street, Seattle, WA 98195, USA.

*To whom correspondence should be addressed. E-mail: quentin.grafton@anu.edu.au

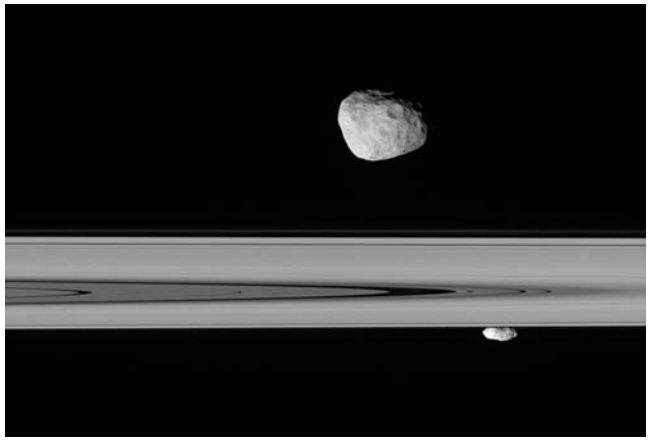
Saturn's Small Inner Satellites: Clues to Their Origins

C. C. Porco,[1]* P. C. Thomas,[2] J. W. Weiss,[1] D. C. Richardson[3]

AUTHORS' SUMMARY

Before Voyager 1 arrived at Saturn in 1980, it was thought that the icy particles composing Saturn's rings were the detritus left over from the planet's formation and that strong tidal forces close to the planet prevented them from aggregating to solids much bigger than a few meters across. The Voyager spacecraft discovered small moons orbiting within and just outside the main rings, and made observations that suggested a ring mass equivalent to that of Saturn's 400-km-wide icy moon Mimas. These and other Voyager findings led many scientists to conclude instead that Saturn's rings formed by the catastrophic collisional disintegration of one or several large icy bodies, perhaps preexisting moons. The resulting debris would have quickly settled into the equatorial plane to form a thin disk. It was possible that the different major rings—A, B, and C—formed from different progenitor bodies and therefore were of different ages.

However, Voyager results also implied that a Mimas-sized moon would survive without collisional disruption for so long that any ring-forming event producing that much debris would have had to happen in the early days of the solar system, ~4.5 billion years ago. Hence, ring B, by far the most massive, might be that old. Ring A, with less mass requiring a smaller, more easily disrupted progenitor body, might have formed more recently (*1*). The presence of moons 30 km in diameter near the outer part of the A ring, such as Pan and Atlas, remained a puzzle: The lifetimes of such small bodies against collision were expected to be shorter still. If disrupted since the A ring's formation, they must have reformed. But how could they do so in the presence of strong planetary tides? It was therefore deemed likely that the small, irregularly shaped moons within and around the rings were shards left over from the original breakup of the rings' progenitor body (or bodies) (*2*). Even the more distant, small moons of Saturn's major satellites Tethys and Dione were expected to be irregular chips off their larger co-orbital companions.



View from the Cassini spacecraft of Saturn's moon Janus (about 190 km across) on the near side of the rings and Prometheus (about 120 km across) on the far side. Both moons have a low density. Analytical results and computer simulations imply that Prometheus, and perhaps Janus, have grown by accumulating porous ring debris around a relic core.

Divining the origin and evolution of Saturn's entire ring-satellite system was one of the principal goals of the Cassini mission. Cassini has surveyed the rings (within the orbit of Titan) for small bodies missed by Voyager and gathered information on their physical characteristics and orbits. Images have yielded sizes and shapes; masses have been determined through observed gravitational effects on other moons or on the rings.

Here, we show that the ring-region moons have high porosities and low densities, about half the density of water ice. Other Cassini results (*3*) imply that the ring particles themselves must have a similar low density. Beyond about the orbit of the F-ring shepherd Pandora, where tidal effects weaken, accretion of such porous material can proceed easily. So whether Janus, Epimetheus, or any of the other, much smaller, more distant moons were formed entirely from aggregation of smaller fragments or are monolithic collisional shards, or somewhere in between is unknown. However, closer to the planet, we find that spontaneous, unseeded aggregation of porous particles, even if broadly distributed in size, to form large, homogeneous kilometer-sized bodies is not feasible.

Instead we suggest, and computer simulations support, that the moons Pan and the newly discovered Daphnis (both within the outer A ring), Atlas (just beyond the A ring), and even Prometheus and Pandora (a few thousand kilometers farther out) all likely grew to their present sizes by the accumulation of porous ring material onto massive, denser cores that were one-third to one-half the present-day sizes of the moons. Although the cores themselves may be collisional shards, the present moons in their entirety cannot be. The cores of Pan and Daphnis were likely large enough to open their respective gaps in the rings, and their growth was largely finished before the gaps were cleared. A secondary stage of accretion apparently formed the equatorial ridges of Pan and Atlas (*4*), and presumably Daphnis as well, after the disk had thinned to its present ~20-m thickness.

How far in the past all this took place is hard to say, and depends on when and whether Pan and Daphnis were disrupted after their initial formation. However, if they suffered any disruptions at all, their cores must have remained more or less intact, or reformation would have been impossible. Perhaps the thick porous blankets of material surrounding these cores today has protected them from being blown to bits, because porous material can absorb impacts with little damage to the underlying structure (*5*). It remains to be tested whether the surfaces of Pan and Daphnis and the other moons in the ring region have the same composition as the rings, which would support this story. Also, crater size-frequency distributions on the larger saturnian satellites, especially Iapetus, will clarify how frequently collisions have occurred within the saturnian system, telling us whether the A ring is likely to be young and its small moons have been blasted apart and reaggregated since their original formation, or whether the whole system has been around since the dawn of the solar system.

Summary References

1. L. Esposito, *Icarus* **67**, 345 (1986).
2. B. A. Smith *et al.*, *Science* **215**, 505 (1982).
3. M. Tiscareno *et al.*, *Astrophys. J.* **651**, L65 (2006).
4. S. Charnoz, A. Brahic, P. C. Thomas, C. C. Porco, *Science* **318**, 1622 (2007).
5. E. Asphaug *et al.*, *Nature* **393**, 437 (1998).

FULL-LENGTH ARTICLE

Cassini images of Saturn's small inner satellites (radii of less than ~100 kilometers) have yielded their sizes, shapes, and in some cases, topographies and mean densities. This information and numerical *N*-body simulations of accretionary growth have provided clues to their internal structures and origins. The innermost ring-region satellites have likely grown to the maximum sizes possible by accreting material around a dense core about one-third to one-half the present size of the moon. The other small satellites outside the ring region either may be close to monolithic collisional shards, modified to varying degrees by accretion, or may have grown by accretion without the aid of a core. We derived viscosity values of 87 and 20 square centimeters per second, respectively, for the ring material surrounding ring-embedded Pan and Daphnis. These moons almost certainly opened their respective gaps and then grew to their present size early on, when the local ring environment was thicker than it is today.

The first close views of the small saturnian ring-region satellites Atlas, Prometheus, Pandora, Janus, and Epimetheus came as a result of the Voyager flybys in the early 1980s (*1*, *2*). Voyager images, generally of spatial resolutions coarser than ~2 km/pixel, yielded approximate sizes and albedos. Estimated normal reflectances of ≥ ~0.5 indicated icy rather than rocky compositions (*3*).

The Voyager-era view of the origin of Saturn's rings (*2*, *4*) called for the catastrophic breakup of one or several large icy bodies within the Roche zone of the planet—the boundary within which like-sized bodies cannot accrete owing to the differential gravity from the planet. However, smaller particles within the Roche zone have no difficulty accreting onto bigger bodies [e.g., (*5*–*7*)]. A progenitor body might have been a recently formed moon that evolved into the Roche zone through aerodynamic drag with the remains of the proto-saturnian nebula before it completely dissipated, or a body that had been captured into orbit around Saturn. The debris from the disruption event(s) became progressively ground down in mutual collisions to create a swarm of particles having a size distribution expected from collisions. The largest collisional shards, which for a system of ice particles would be roughly ~10 km in diameter (*4*), presumably were spared this fate by opening gaps in the rings and terminating their own erosion, assuming that the velocity dispersion of the particles had reduced the ring thickness to a size comparable to the size of the shards.

Other small moons throughout the Saturn system were also generally thought to be collisional shards (*2*). Janus and Epimetheus, occupying as they do essentially the same orbit, were thought to be the two major remnants of a catastrophic disruption of a single progenitor body. Farther from the planet, the small Trojans of Dione and Tethys (Hellene, Calypso, and Telesto) were believed to be chips off their parent moons, Dione and Tethys, or even off the other large saturnian moons, that were then caught into resonance with Dione and Tethys.

The Imaging Science experiment (*8*) on the Cassini spacecraft has imaged these eight bodies, plus Pan in the Encke gap of the A ring, at image scales better than 1 km/pixel (Figs. 1 and 2). It has discovered three very small moons—Methone, Anthe, and Polydeuces—and has recovered one, Pallene, previously discovered by Voyager, all within the main satellite system; it has also found Daphnis within the Keeler gap of the outer A ring (*9*–*11*). Cassini has also had a close encounter with one of them, Polydeuces. Image analyses have yielded sizes, shapes, topography, and in some cases masses and, hence, densities.

With this information in hand, we have examined these bodies for similarities and trends in physical characteristics, and compared them to numerical simulations of accretion from a system of particles.

Data and reduction methods. The image analysis technique chosen to measure the size and shape of a given satellite depends on the spatial resolution and coverage of a particular body (*12*–*15*). The moons' measured and modeled physical characteristics (Table 1) include the semi-axes of an ellipsoid that provides the best fit to the shape. (The observed shape is the collection of measured {*x*,*y*,*z*} coordinates of each grid point on a 5° × 5° latitude-longitude grid on the moon's surface.) These semi-axes may give slightly different volumes than those derived directly from the shapes because of imprecision in fitting simple analytic functions to complex natural forms. The moons' mean densities were computed using the actual observed volumes and the best measurements of masses for those moons for which masses have been determined (*16*).

The topography of those moons for which we have reliable shapes, masses, and spin states has also been calculated (Fig. 3). [For completeness, we include topographic models for those moons imaged at sufficient resolution and coverage; we assume an internal density where masses are unavailable (figs. S1 to S3).] Pan and Atlas are distinguished from the rest by distinct, longitudinally asymmetric, equatorial ridges delineated by concavities in the moons' limb profiles (Figs. 2 and 3).

Analysis. We compared the observed satellite properties with those expected for gravitationally bound agglomerations of particles ("rubble piles") that have formed by accretion. Such loose accumulations might be expected to have the shape of the Roche lobe (*5*, *17*). Material can no longer accrete onto a body because of gravity alone once it reaches the size of its Roche lobe. We define the critical density as the density of a body that entirely fills its Roche lobe:

$$\rho_{crit} = \frac{3M_p}{\gamma a_{orbital}^3} \quad (1)$$

where M_p is the mass of the planet, $a_{orbital}$ is the semimajor axis of the orbit, and γ is a dimensionless shape parameter so that the volume of the moonlet is γa^3, where a is the long semi-axis. For the Roche lobe,

$$\gamma = \frac{2\pi \ln(2+\sqrt{3})}{3\sqrt{3}} \approx 1.59 \quad (2)$$

For a sphere, $\gamma = 4\pi/3$. Roche lobe dimensions for each moon whose mass is known are given in Table 1. The long semi-axis, a_R, points toward the planet, and the "perpendicular" axis, c_R, lies perpendicular to the orbit plane. The axis ratios for the Roche lobe are a_R:b_R:c_R = 3:2:2. For the computation of the critical density, we use the volume of a Roche lobe appropriate for a point-mass source.

Table 1 reveals that the observed long axes of Pan, Daphnis, Atlas, and Prometheus are within 15% of the long axis of the Roche lobe for a body of the given mass at the satellites' orbit. The trend breaks with Janus and Epimetheus, the co-orbital moons, which are much more discrepant in size. The inner satellites have surprisingly low densities: ~0.4 to 0.6 g cm^{-3} (Table 1). The densities of the innermost of these moons are all approximately equal to the critical density at that distance, whereas farther from Saturn, the densities of Janus and Epimetheus are not. These comparisons immediately suggest a formation scenario (for at least the innermost moons) of growth by accretion of material until the critical density is reached, at which point the moon more or less fills its Roche lobe.

Numerical simulations. We have examined accretion into a "rubble pile" with the use of a numerical code that realistically simulates the behavior of ring particles in a patch of fixed dimensions in orbit around Saturn, inclusive of collisional and self-gravitational effects (*18*, *19*). In this patch, we embedded, in turn, solid cores with radii of 12.5 and 25 m and internal densities of 0.9 g cm^{-3} appropriate for solid water ice. We chose particle disk characteristics—optical depth and surface mass density—believed to be representative of Saturn's A ring (*20*, *21*). The ring particles were given a differential power-law size distribution, $dN(r) \propto r^{-q}dr$, with $q \sim 3$ and radii r ranging from 0.3 to 5 m (*22*), a mass density of 0.4 g cm^{-3} reasonable for porous icy bodies, and elastic properties expected for ice (*23*). The size of the patch was chosen to be at least 10 times the final size of the moonlet's Roche lobe to minimize interactions between the moonlets and their mirror images in

[1]Cassini Imaging Central Laboratory for Operations (CICLOPS), Space Science Institute, 4750 Walnut Street, Boulder, CO 80301, USA. [2]Center for Radiophysics and Space Research, Cornell University, Ithaca, NY 14853, USA. [3]Department of Astronomy, University of Maryland, College Park, MD 20742, USA.

*To whom correspondence should be addressed. E-mail: carolyn@ciclops.org

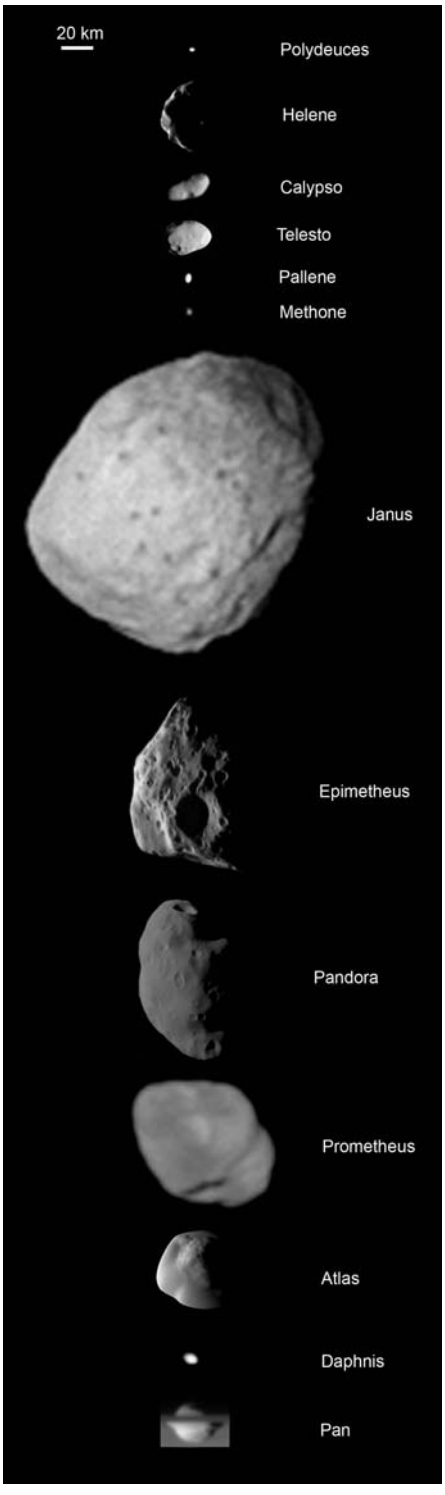


Fig. 1. The highest-resolution images of all the small saturnian moons, except Anthe (between Methone and Pallene), shown to the same scale and in order of distance from Saturn. Pan is transiting Saturn and is partially occulted by the edge-on rings.

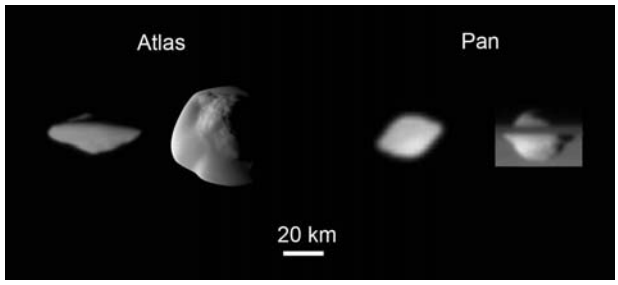


Fig. 2. The highest-resolution images of Pan and Atlas, showing their distinctive "flying saucer" shapes, owing to equatorial ridges not seen on the other moons. From left to right: a view of Atlas' trailing hemisphere, with north up, at a spatial scale of ~1 km/pixel; Atlas seen at ~250 m/pixel from mid-southern latitudes, with sub-Saturn hemisphere pointing up and leading hemisphere to the left; Pan's trailing hemisphere seen at ~3 km/pixel from low southern latitudes; an equatorial view, with Saturn in the background, of Pan's anti-Saturn hemisphere at ~1 km/pixel. On Atlas, the ridge extends 20° to 30° in latitude on either side of the equator; on Pan, its latitudinal extent is 15° to 20°. Atlas shows more asymmetry than Pan in having a more rounded ridge in the leading and sub-Saturn quadrants. The heights of the ridges can be crudely estimated by assuming ellipsoidal shapes that lack ridges and vary smoothly across the equator (fig. S4). Heights of Atlas' ridge range from ~3 km (270°W) to 5 km (180° and 0°W); Pan's ridge reaches ~4 km at 0°W and is ~1.5 km over most of the rest of the equator. The ridge represents ~27% of Atlas' volume and 10% of Pan's volume.

neighboring virtual cells (which are needed to provide the boundary conditions at the patch edges).

As expected, the core accreted particles until a statistical equilibrium was reached—as many particles were accreted as were dislodged by incoming new particles—and its effective bulk density dropped to the critical density (i.e., the value where it exactly fills its Roche lobe). The a, b, and c axes of the simulated moonlet were in a ratio of approximately 6:4:3, close to the effective Roche lobe axes ratio of 6:4:4 (Fig. 4). The discrepancy may be due to finite particle size effects in our simulations, where some of the accreted particles are smaller than the core by factors of only 2 to 5. An alternative explanation is that the particles with low relative velocity tend to enter through the inner and outer Lagrange points at the moonlet's equator and then migrate to different longitudes as well as higher latitudes, and this process may not be 100% effective at filling the Roche lobe in the perpendicular direction. The low bulk density and implied global porosities of the moonlet (~30 to 65%) are consistent with porosities inferred for many asteroids, commonly 20 to 60% (*24*), and for icy Hyperion, ~43% (*25*).

Discussion. The expected final size of a growing moon with a long semi-axis a, relative to the initial core with radius R_c, can be calculated as

$$\frac{a}{R_c} = \left(\frac{4\pi}{3\gamma}\frac{\rho_c - \rho_m}{\rho_{crit} - \rho_m}\right)^{1/3} \quad (3)$$

where ρ_{crit} is the critical density, ρ_c is the density of the core, and ρ_m is the density of the mantle of material accreted around the core. (Note that ρ_m is equal to the bulk density of the particles, ρ_p, times their filling factor in the mantle; this equation is applicable only when $\rho_m < \rho_{crit}$.) We define mantle "macroporosity" to be $1 - (\rho_m/\rho_p)$, a measure of the degree of packing of the accreted material. The macroporosity of the accreted material, with a bulk density of $\rho_p = 0.4$ g cm^{-3} in our simulations, is ~65%.

Particles with bulk densities very near or below the critical density of 0.4 g cm^{-3} will not accrete material, and homogeneous rubble piles with no central mass concentration are not even stable against tidal shear in the ring region—hence the need for a core of larger density to begin the process of accretionary growth to a stable moon. If the core is a kilometer-sized collisional shard resulting from the breakup of a bigger icy progenitor ring body (*4*), it is reasonable to assume that its density might be somewhere between 0.5 and 0.9 g cm^{-3}. We find that for this plausible range of core densities, the ring-region moons had to begin with cores that were approximately one-half to one-third the size of the final body (Fig. 5). In that case, neither ~40-km-radius Prometheus nor ~4-km-radius Daphnis could have begun with a 10-m-radius core.

Despite the gross difference in the core sizes compared (i.e., 12.5 m versus tens of kilometers), the ratio of the escape speed from the moonlet to the tidal shear across it, $V_{esc}/\Omega R$, is a constant for the same mean density, independent of moonlet size. ($V_{esc}/\Omega R$ is a measure of the strength of the moon's gravity pulling on a particle relative to the tidal forces that tend to keep particles from accreting.) This is also suggested by the identical results we achieved when we doubled the core size in our simulations (Fig. 5). Thus, for the purposes of investigating the general character of accretion, our numerical results are essentially scale-invariant and should apply to much larger bodies at the same orbital distance from Saturn, although nonspherical ring particle shapes and finite size effects may alter the mantle porosity somewhat from small to large sizes.

The observed moons deviate from exact Roche lobe shapes. Indeed, almost all of them have intermediate axes larger than those of their Roche lobes (Table 1). For Pan and Atlas, their shapes are more like flying saucers (Fig. 2), and even the ridgeless forms of Pan and Atlas show a departure from the classic Roche shape (fig. S4). The point-source assumption in computing the Roche lobe shape, which might be expected to contribute to these discrepancies, in fact contributes very little (*26*). Other processes, however, can affect the detailed shapes of an accreting satellite. Those that work to make the surface approximately an equipotential, such as fluid behavior or downslope transport of loose material, generally fail for objects less than 100 km in radius (*27*). Departures from a pure Roche lobe shape might instead be expected as a result of the details of accretion. For example, particles carrying sufficient kinetic energy can strike

Table 1. Characteristics of Saturn's small moons. Orbital distances from Saturn, $a_{orbital}$, are from (*29*) for all moons except Anthe (*11*), Telesto, Calypso, and Helene; orbital distances of the latter three are taken from the JPL ephemeris file SAT252. Masses are determined from orbital integrations [i.e., satellite-satellite perturbations (*27*)] or, for Pan (*9*) and Daphnis (*32*), from the effects of the moons on the rings. Mean radius, r_m, is that for a sphere having the same volume as the observed moon, except for Anthe, whose mean radius has been estimated from its brightness (*11*). Axes *a*, *b*, and *c* are those of the best-fit model ellipsoid; a_R, b_R, and c_R are the Roche lobe axes, computed using the moons' masses and assuming a point-mass source. Critical densities, ρ_{crit}, are computed as described in the text. Moons' estimated densities, ρ, are obtained from satellite masses and observed shapes. Fitted ellipsoid dimensions for Pan and Atlas with ridges removed are 15.9, 15.1, and 10.7 km, and 16.9, 16.0, and 9.3 km, respectively.

Satellite	$a_{orbital}$ (km)	Mass ($\times 10^{19}$ g)	r_m (km)	a (km)	b (km)	c (km)	$\pm \delta a, \delta b, \delta c$ (km)	c/a	a_R (km)	$b_R = c_R$ (km)	a/a_R	ρ_{crit} (g cm^{-3})	ρ (g cm^{-3})	ρ/ρ_{crit}
Pan	133,584	0.495 ± 0.075	14.2 ± 1.3	17.4	15.8	10.4	2.0, 1.3, 0.84	0.60	19.1	12.7	0.91	0.45	0.41 ± 0.15	0.92 ± 0.32
Daphnis	136,504	0.0084 ± 0.0012	3.9 ± 0.8	4.5	4.3	3.1	0.8, 0.8, 0.9	0.69	4.9	3.2	0.93	0.42	0.34 ± 0.21	0.80 ± 0.50
Atlas	137,670	0.66 ± 0.06	15.1 ± 1.4	20.9	18.1	8.9	1.4, 2.5, 0.8	0.43	21.6	14.4	0.97	0.41	0.46 ± 0.10	1.12 ± 0.24
Prometheus	139,380	15.67 ± 0.20	43.1 ± 2.7	66.3	39.5	30.7	3.2, 3.2, 2.0	0.46	62.9	41.9	1.05	0.40	0.47 ± 0.065	1.18 ± 0.17
Pandora	141,720	13.56 ± 0.23	40.3 ± 2.2	51.6	39.8	32.0	1.8, 2.1, 2.9	0.62	61.0	40.7	0.85	0.38	0.50 ± 0.085	1.32 ± 0.23
Epimetheus	151,410	53.07 ± 0.14	56.7 ± 1.9	58.0	58.7	53.2	2.5, 3.2, 0.8	0.92	102.7	68.4	0.57	0.31	0.69 ± 0.13	2.25 ± 0.42
Janus	151,460	188.91 ± 0.50	89.6 ± 2.0	97.4	96.9	76.2	2.9, 2.2, 1.2	0.78	157.4	105	0.62	0.31	0.63 ± 0.063	2.03 ± 0.21
Methone	194,440	–	1.6 ± 0.6	–	–	–	0.6	–	–	–	–	0.15	–	–
Anthe	197,700	–	~1	–	–	–	–	–	–	–	–	0.14	–	–
Pallene	212,280	–	2.2 ± 0.3	2.6	2.2	1.8	0.4, 0.3, 0.2	0.69	–	–	–	0.11	–	–
Telesto	294,710	–	12.4 ± 0.4	15.7	11.7	10.4	0.6, 0.3, 0.3	0.66	–	–	–	0.04	–	–
Calypso	294,710	–	10.6 ± 0.7	15.0	11.5	7	0.3, 2.3, 0.6	0.47	–	–	–	0.04	–	–
Polydeuces	377,200	–	1.3 ± 0.4	1.5	1.2	1.0	0.6, 0.4, 0.2	0.67	–	–	–	0.02	–	–
Helene	377,420	–	16.5 ± 0.6	19.4	18.5	12.3	0.2, 1.0, 1.0	0.63	–	–	–	0.02	–	–

a surface and stick because of cohesive and frictional forces, even if the impact site has a considerably higher gravitational potential than other locales on the surface. Alternatively, hypervelocity impacts may form large craters, redistribute material around the surface (or eject it entirely), and in general compete with low-speed accretion and consequently lead to deviation from an equipotential surface. Accretion of particles with low relative velocity that are essentially coplanar with the satellite and are therefore on low-inclination orbits, might also build up the moon's intermediate axes long before the Roche lobe has been "filled." Finally, a moon's detailed orbital motion can determine where around the midsection the material will ultimately land. The equatorial ridges of Pan and Atlas and their specific longitudinal asymmetries can be explained by a post-formation, late-stage accretion (*28*) from a vertically thin disk. The details depend on the eccentricities and inclinations of these two moons (*29*). (Daphnis too may sport a ridge, owing to accretion from the thin rings, that Cassini may reveal in future observations.) Thus, it is not surprising that the moons do not have exact Roche lobe shapes.

We tested the degree to which interior models with dense cores move the surfaces toward equipotentials by calculating (*12*) the surface potentials and the effective gravitational topography on the four ring moons, using their observed shapes and assuming a central core of 0.9 g cm^{-3} and a mantle of density 0.15 g cm^{-3} (Fig. 3). The gravitational topography on the ring-region moons Pan, Atlas, Prometheus, and Pandora reaches a few kilometers, or modest fractions of the mean radii. The very low values of gravity on the Saturn and anti-Saturn equatorial regions indicate that these objects are in fact near the limit of accretion in those locales, as expected. These features are not unique to "core" interior models: Assuming a uniform-density interior produces comparable gravitational topography (Fig. 6). However, because a dense core is required to start the accretion process, the observations, core models, and simulations are consistent with a formation scenario, for at least the innermost moons, of growth by accretion around a dense core to the maximum possible size (and minimum critical density), wherein the overall Roche lobe conditions and the varied dynamics of impacting particles dominate any possible effects due to surface processes such as downslope motion.

If this accretionary growth model is correct, we can now add an additional factor to the dynamic history of ring-embedded moons such as Pan and Daphnis, which currently maintain gaps—Encke and Keeler, respectively—in Saturn's outer rings. The mass of a ring-embedded moonlet required to open and maintain a narrow gap at a given orbital radius can be found by balancing the viscous ring torque with the torque exerted by a satellite on the gap edge (*30*):

$$\mu^2 = \frac{3\pi\nu}{0.84(a_{orbital})^2\Omega}\left(\frac{w}{a_{orbital}}\right)^3 \qquad (4)$$

where Ω is the orbital frequency at semimajor axis $a_{orbital}$, μ is the ratio of the mass of the moonlet to the mass of Saturn, w is the half-width of the gap, and ν is the viscosity of the ring. For Pan and Daphnis to maintain their present gaps, the ring viscosities must be $\nu \sim 87$ cm^2 s^{-1} for Pan and $\nu \sim 20$ cm^2 s^{-1} for Daphnis. Using these viscosities, we can find the minimum mass required to open the Encke and Keeler gaps. That point is reached when the moon grows to a mass where the gap width, maintained by its overlapping Lindblad resonances at the gap's edge, exceeds the moon's Roche lobe (assuming that the ring's thickness is not much larger than the moon).

Replacing w with the Roche lobe radius $[a_{orbital}(\mu/3)^{1/3}]$, we obtain

$$\mu = \frac{\pi\nu}{0.84(a_{orbital})^2\Omega} \qquad (5)$$

Using the viscosities derived above produces a minimum Pan mass for opening the Encke gap of $GM_P \sim 5.5 \times 10^{-7}$ km^3 s^{-2}, yielding a Roche lobe semi-axis $a \sim 2$ km; for Daphnis and the Keeler gap, $GM_D \sim 1.2 \times 10^{-7}$ km^3 s^{-2} and $a \sim 1$ km. Daphnis' mass today is 40 times this minimum for the Keeler gap; Pan's mass is 600 times the minimum to open the Encke gap. Their original cores must have been between 1.8 and 3 km in radius for Daphnis, and ~8 to 12 km for Pan, depending on the core density (Fig. 5). It is likely that both bodies, but especially Pan, opened their gaps before growing to their final sizes (and dropping to their critical density).

In order for Pan and Daphnis to grow from small cores to the sizes they are today, both the initial gap-opening and accretionary epochs for these moons must have occurred during a time when the local debris disk they were embedded in was at least comparable to the final size of the moon and not as vertically thin (~20 m) as Saturn's present rings—that is, very early in the history of the outer A ring. (By analogy, the same would be true even for Atlas, Prometheus, and Pandora.) Because times for collapse to a thin disk are short (*4*), the accretion stage must also have occurred rapidly. A later secondary accretionary stage for Pan and Atlas, after the local disk had thinned to its present state but was still amply populated, was likely responsible for adorning these two moons with the distinct equatorial ridges we see today (*28*).

The possibility that all these ring-region moons were catastrophically disrupted and reaccreted, perhaps several times, during the lifetime of the A ring (*2*) does not fundamentally alter the above

sequence of events, but only resets the time of last occurrence. This assumes that the dense cores more or less survived the disruption and were therefore available to initiate the next accretionary epoch. The thick shells of ring material surrounding the cores may in fact have protected them from breakup (*31*). In this case, the debris clouds from which today's observed moons coalesced would simply consist of material recycled from their most recent disruption.

The discrepancy between the Roche lobe size and the sizes of the more distant moons, such as Janus and Epimetheus, may imply that these moons ran out of ring material before reaching their critical densities and maximum sizes. These moons very likely originated closer to Saturn and, through the torques they exert on the rings, quickly moved out to their present position (*30*). Taking their possible radial migration into consideration, their sizes are still significantly smaller, and their mass densities significantly higher, than one would expect for a Roche lobe–filling body in their general vicinity. If their formation location was more distant than ~148,000 km from Saturn, it is possible that they are entirely accretional in origin, as no core is required for accretion beyond this distance assuming the ring particle mass densities we have adopted here (Fig. 5). However, if they formed within an orbital distance of ~148,000 km from Saturn, they may in fact be the original shards, perhaps modified to some degree by accretion, that resulted from the disruption of their progenitor body (*2*).

Within the observational uncertainties, the tiny satellite-region moons beyond Janus (i.e., Methone, Anthe, Pallene, and Polydeuces) could be any shape. Of the larger Trojans, only Telesto's shape looks Roche-like, and high-resolution images show it to be visibly coated with debris, implying perhaps some accretion. The critical densities at the orbital distances of all these moons are so small that it is exceedingly doubtful that these moons are filling their Roche lobes; such densities imply implausibly high macroporosities approaching ~95%. All these moons, like Janus and Epimetheus, may be original collisional shards that have been modified by accretion to varying degrees by some combination of the processes mentioned above, or they may have formed entirely by accretion without the aid of a dense core.

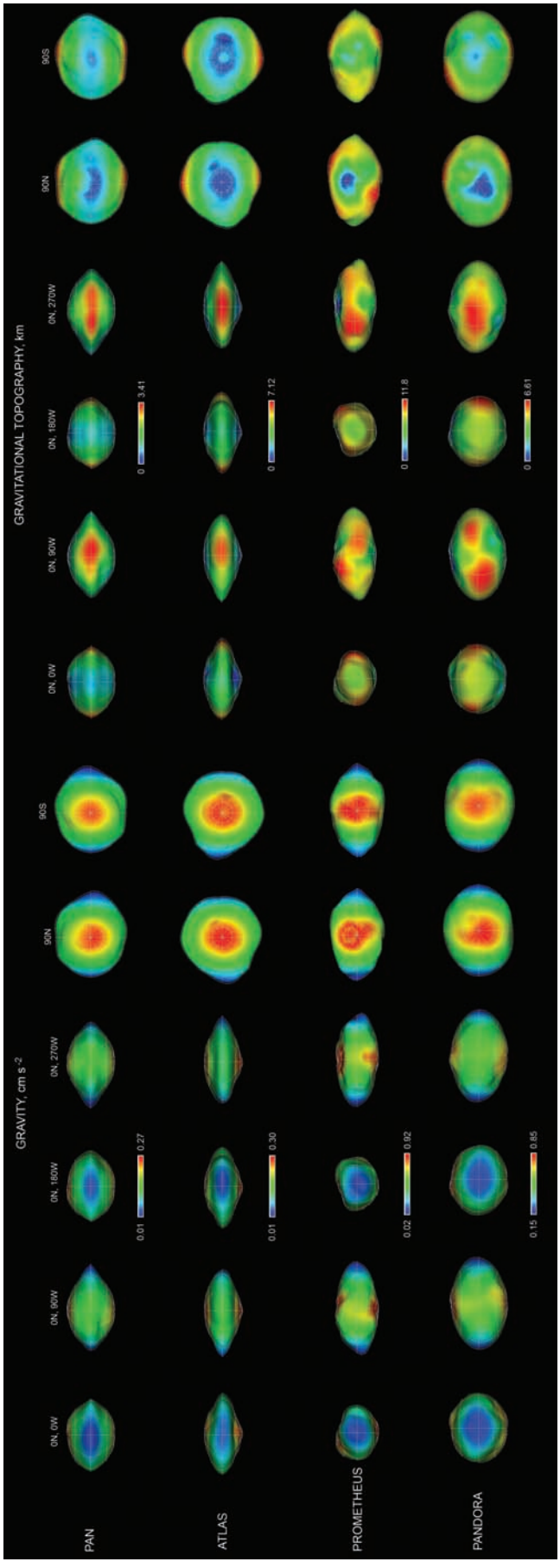


Fig. 3. Modeled gravitational topography for the ring-region moons Pan, Atlas, Prometheus, and Pandora, assuming core and mantle densities of 0.9 g cm^{-3} and 0.15 g cm^{-3}, respectively. The figures are based on the appropriate spin, tidal forces, and potential energy at the surface. Left panels show magnitude of surface accelerations. Right panels show gravitational topography, which is the potential energy (relative to the south pole) divided by an average acceleration. Views, from left to right, are of the Saturn-facing, leading, anti-Saturn, trailing, northern, and southern hemispheres.

References and Notes

1. B. A. Smith *et al.*, *Science* **212**, 163 (1981).
2. B. A. Smith *et al.*, *Science* **215**, 505 (1982).
3. P. C. Thomas, J. Veverka, D. Morrison, M. Davies, T. V. Johnson, *J. Geophys. Res.* **88**, 8743 (1983).
4. A. Harris, in *Planetary Rings*, R. Greenberg, A. Brahic, Eds. (Univ. of Arizona Press, Tucson, AZ, 1984), pp. 641–659.
5. D. R. Davis, S. J. Weidenschilling, C. R. Chapman, R. Greenberg, *Science* **224**, 744 (1984).
6. A. R. Dobrovolskis, J. A. Burns, *Icarus* **42**, 422 (1980).
7. R. M. Canup, L. W. Esposito, *Icarus* **113**, 331 (1995).
8. C. C. Porco *et al.*, *Space Sci. Rev.* **115**, 363 (2004).
9. C. C. Porco *et al.*, *Science* **307**, 1226 (2005).
10. C. C. Porco *et al.*, *IAU Circ. 8432* (2004).
11. C. C. Porco *et al.*, *IAU Circ. 8857* (2007).
12. P. C. Thomas, *Icarus* **105**, 326 (1993).
13. D. P. Simonelli, P. Thomas, B. T. Carcich, J. Veverka, *Icarus* **103**, 49 (1993).

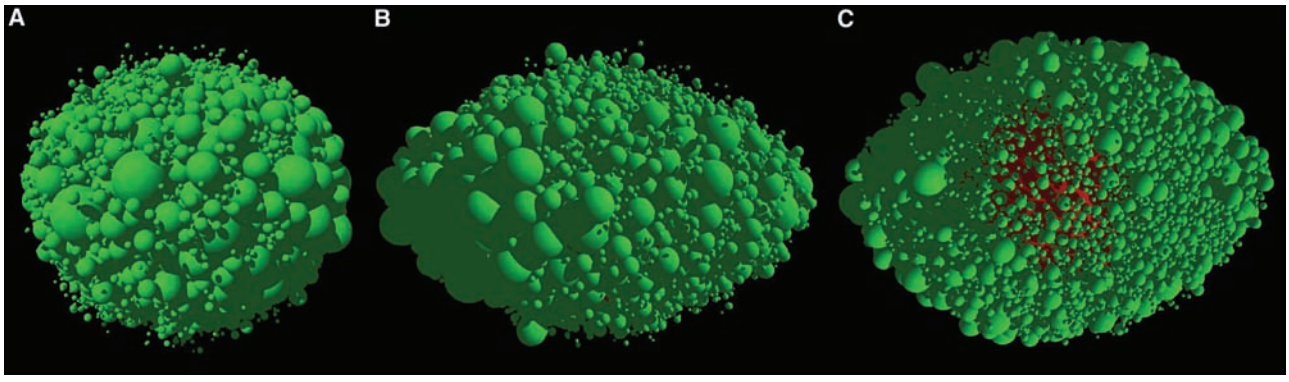


Fig. 4. Ring-patch code simulations with an embedded moonlet of radius 25 m. (**A**) View of the Saturn-facing hemisphere. (**B**) View of the leading hemisphere. (**C**) View from the northern hemisphere, with the leading hemisphere pointing downward. The red region visible in these images is the core, slightly visible in (C). The observed axes ratio in these simulations was 6:4:3, slightly smaller than the classic Roche lobe shape of 6:4:4 in the vertical direction.

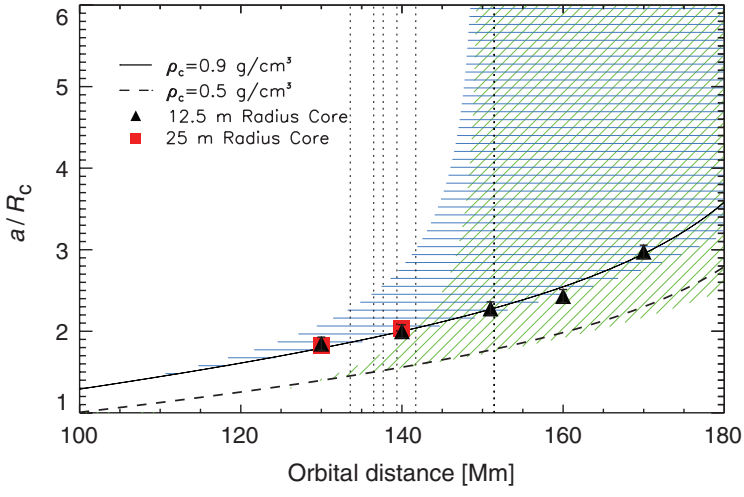


Fig. 5. Variation with orbital radius from Saturn of the final relative size of a moon (with respect to the size of the core) growing in a disk of particles with internal bulk mass density of 0.4 g cm^{-3}. Green hatched region is the range of final sizes for a moon with a core of 0.5 g cm^{-3}; blue hatched region, for a moon with a core of 0.9 g cm^{-3}. Both are bounded by mantle porosities of 20% (top) and 70% (bottom), or mantle densities of 0.32 g cm^{-3} and 0.12 g cm^{-3}, respectively. Data points are the simulation results for a core with a radius of 12.5 m (triangles) and 25 m (red squares), a density of 0.9 g cm^{-3}, and a porosity of 65% (or a mantle density of 0.15 g cm^{-3}). Error bars on the simulations are comparable to the size of the symbols and show the statistical variations in the final moonlet size over the last five orbits of the 15-orbit simulations. The solid and dashed black curves are the theoretical predictions for moons with two different core densities and the mantle porosity (~65%) found in all the simulations. The agreement between data and theory confirms the relation used to calculate the hatched regions. Vertical dotted lines signify the locations of the moons considered in this paper: from left to right, Pan, Daphnis, Atlas, Prometheus, Pandora, and Janus/Epimetheus. Beyond ~148,000 km, moonlet growth does not require a dense core; accretion and growth from a debris disk can proceed unaided.

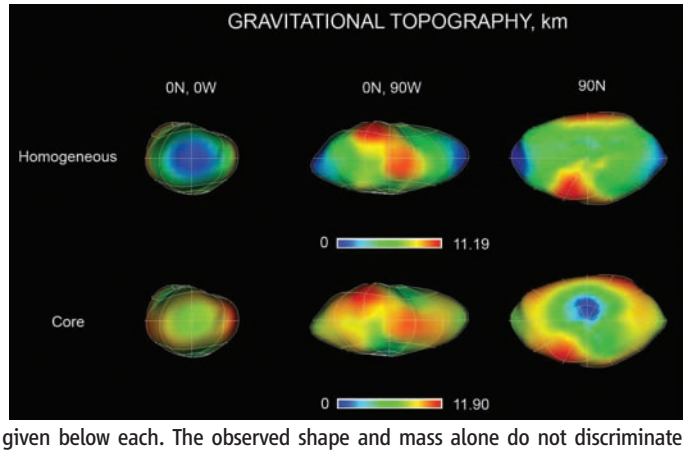


Fig. 6. Gravitational topography for Prometheus, assuming its observed shape but two different interior models: homogeneous and core. The former assumes a uniform interior with critical density for Prometheus of 0.4 g cm^{-3}; the latter assumes a core of 0.9 g cm^{-3} and mantle of 0.15 g cm^{-3}, with an overall density equal to its critical density. The range of gravitational topography in kilometers is given below each. The observed shape and mass alone do not discriminate between these two interior models.

14. P. C. Thomas *et al.*, *Icarus* **135**, 175 (1998).
15. See supporting material on *Science* Online.
16. Masses are unavailable for the Trojans of Tethys (Calypso and Telesto) and Dione (Helene and Polydeuces), as well as the tiny satellite-region moons Methone, Anthe, Pallene, and Polydeuces, because none of these is massive enough to measurably perturb any nearby or resonant moons.
17. J. M. A. Danby, *Fundamentals of Celestial Mechanics* (Willmann-Bell, Richmond, VA, 1992).
18. D. C. Richardson, *Mon. Not. R. Astron. Soc.* **269**, 493 (1994).
19. D. C. Richardson, T. Quinn, J. Stadel, G. Lake, *Icarus* **143**, 45 (2000).
20. J. Colwell, L. W. Esposito, M. Sremcevic, *Geophys. Res. Lett.* **33**, L07201 (2006).
21. M. S. Tiscareno *et al.*, *Nature* **440**, 648 (2006).
22. R. G. French, P. D. Nicholson, *Icarus* **145**, 502 (2000).
23. K. D. Supulver, F. G. Bridges, D. N. C. Lin, *Icarus* **113**, 188 (1995).
24. D. T. Britt, D. Yeomans, K. Housen, G. Consolmagno, in *Asteroids III*, W. F. Bottke, A. Cellino, P. Paolicchi, R. P. Binzel, Eds. (Univ. of Arizona Press, Tucson, AZ, 2002), pp. 485–500.
25. P. C. Thomas *et al.*, *Nature* **448**, 50 (2007).
26. It is reasonable to question the assumption that the moonlets are point masses in calculating the Roche lobe dimensions (shape and critical density) because they are clearly extended bodies. However, no self-gravitating extended body with a homogeneous interior and the low densities found for the moons in this study is even stable in the ring region, and bodies with dense cores and under-dense mantles differ in their dimensions from those of a point-mass Roche lobe by at most ~5%.
27. I. Sharma, J. T. Jenkins, J. A. Burns, *Icarus* **183**, 312 (2006).
28. S. Charnoz, A. Brahic, P. C. Thomas, C. C. Porco, *Science* **318**, 1622 (2007).
29. J. Spitale *et al.*, *Astron. J.* **132**, 692 (2006).
30. P. Goldreich, S. Tremaine, *Annu. Rev. Astron. Astrophys.* **20**, 249 (1982).
31. E. Asphaug *et al.*, *Nature* **393**, 437 (1998).
32. J. W. Weiss, C. C. Porco, M. S. Tiscareno, J. A. Burns, L. Dones, *Bull. Am. Astron. Soc.* **37**, 767 (2005).

33. We thank the staff members of CICLOPS and B. Carcich (Cornell University) for technical support, and L. Dones for fruitful discussions. Supported by NASA Planetary Geology and Geophysics grant NAG5-4451 and the Cassini Project.

Supporting Online Material
www.sciencemag.org/cgi/content/full/318/5856/1602/DC1
Materials and Methods
Figs. S1 to S4

18 April 2007; accepted 5 October 2007
10.1126/science.1143977

Structure of a Site-2 Protease Family Intramembrane Metalloprotease

Liang Feng,[1]* Hanchi Yan,[1]* Zhuoru Wu,[1]* Nieng Yan,[1] Zhe Wang,[2] Philip D. Jeffrey,[1] Yigong Shi[1]†

AUTHORS' SUMMARY

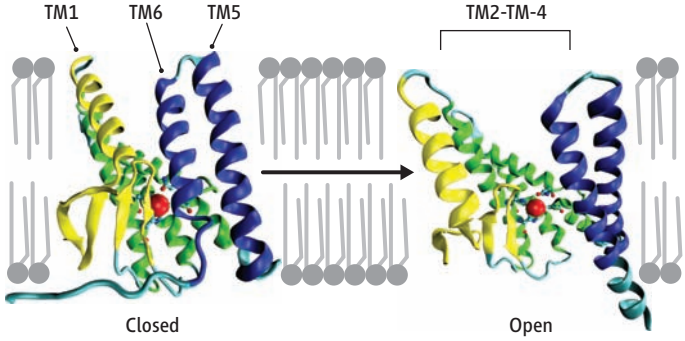
An unusual signaling mechanism involves the cleavage of a transmembrane protein within the lipid membrane bilayer. Cleavage is accomplished by a membrane-embedded protease enzyme (*1*), so this process is called regulated intramembrane proteolysis (RIP) and is used by organisms from bacteria to humans. The first RIP system to be molecularly characterized was the cleavage of the membrane-anchored transcription factor SREBP (sterol regulatory element–binding protein) by a metalloprotease known as site-2 protease (S2P) (*2*). This cleavage releases a transcription factor, which translocates into the nucleus of the cell and activates genes involved in synthesis and uptake of cholesterol and fatty acids. How S2P cleaves a protein embedded within the lipid bilayer has been enigmatic. Cleavage of a protein requires water molecules. How does water gain access to the buried active site of S2P? Is the metal ion cofactor that is required for catalysis exposed to the lipid environment? If not, how can a substrate protein get into the active site? Finally, what does an S2P protease look like, and how does the structure support its function? Our study of the S2P protein provides clues for the answers to all these questions.

X-ray crystallography is a powerful approach for elucidation of the detailed three-dimensional structure of macromolecules. A prerequisite is generation of crystals that are sufficiently ordered and large enough to produce useful x-ray diffraction data—a daunting challenge for membrane proteins. We were able to crystallize the S2P protein from the archaebacterial species *Methanocaldococcus jannaschii* by including the detergent decyl-β-D-maltopyranoside. To improve the ability of these crystals to diffract x-rays, we incorporated two additional detergents into the crystallization buffer. The structure was determined with multi-wavelength anomalous dispersion.

S2P has six transmembrane segments, TM1 through TM6 (see the figure). The catalytic zinc atom is located ~14 Å from the lipid membrane surface. Zinc is coordinated by three amino acids, His^{54} and His^{58} in TM2, and Asp^{148} in TM3, which are highly conserved in all S2P proteins. TM2 and TM4 are stabilized by TM3, and together, these three segments constitute a core domain of S2P (green). Amino acid sequences for TM2–4 are similar in S2P proteins from other species, which suggests that they have a similar structure and a conserved active-site conformation. In the crystals, two molecules of S2P are contained in one asymmetric unit, the minimal element that can be built into an entire crystal. The two molecules exist in different conformations (see the figure). Although the conformations of the core domain are identical, the other TM segments are quite different; TM1 and TM6 are 10 to 12 Å farther apart in one S2P molecule than in the other. The conformational difference has a direct consequence: The active site is accessible only in the S2P molecule in which TM1 and TM6 are farther apart. Hence, these two conformations likely represent the open and closed states of S2P. In the open state, the cleft between TM1 and TM6 can accommodate a peptide in an extended conformation. Thus, we propose that, to be properly positioned for cleavage, the peptide gains access to the active site of S2P through the lateral movement of TM1 and TM6. In the closed state, water molecules can get to the zinc in the active site through a hydrophilic channel that opens to the cytoplasmic side of the lipid membrane.

Although our structure suggests how the substrate peptide may get to the active site, further insight must await biochemical experiments. Given the flexibility of TM1, TM5, and TM6, can substrate enter the active site between TM1 and TM2 or between TM6 and the core domain? Although these possibilities cannot be ruled out, they are not supported by available sequence or structural information. Nonetheless, crystals of S2P were generated in the presence of detergents, rather than membrane lipid. Consequently, the influence of detergents on the structure remains to be characterized.

Despite these caveats, the structure of S2P serves as a framework for understanding the function of intramembrane metalloproteases. The sequence conservation among S2P family members suggests that the active site is in a similar position throughout the family. In contrast, substrate proteins are cleaved at different positions along their putative transmembrane helices. This suggests that, before cleavage, SP2 must recognize a specific sequence in the substrate to appropriately position the cleavage site. Such recognition does not necessarily occur within the lipid bilayer, as is the case for human S2P and its homolog in *Bacillus subtilis*.



Closed and open conformations of an S2P metalloprotease, which cleaves its protein substrates within the cell membrane. Substrate peptide is proposed to gain access to the catalytic zinc atom (red sphere) only in the open conformation.

In addition to the S2P family, there are three additional families of intramembrane proteases: serine protease rhomboid, aspartate protease presenilin, and signal peptide peptidase. The mechanisms of water entry and substrate access appear to be similar between S2P and rhomboid, the only other intramembrane protease for which structural information is available. It remains to be seen whether such mechanisms also apply to the aspartate proteases.

Summary References

1. M. S. Wolfe, R. Kopan, *Science* **305**, 1119 (2004).
2. M. S. Brown, J. Ye, R. B. Rawson, J. L. Goldstein, *Cell* **100**, 391 (2000).

CREDIT : CHRIS BICKEL/*SCIENCE*

FULL-LENGTH ARTICLE

Regulated intramembrane proteolysis by members of the site-2 protease (S2P) family is an important signaling mechanism conserved from bacteria to humans. Here we report the crystal structure of the transmembrane core domain of an S2P metalloprotease from *Methanocaldococcus jannaschii*. The protease consists of six transmembrane segments, with the catalytic zinc atom coordinated by two histidine residues and one aspartate residue ~14 angstroms into the lipid membrane surface. The protease exhibits two distinct conformations in the crystals. In the closed conformation, the active site is surrounded by transmembrane helices and is impermeable to substrate peptide; water molecules gain access to zinc through a polar, central channel that opens to the cytosolic side. In the open conformation, transmembrane helices α1 and α6 separate from each other by 10 to 12 angstroms, exposing the active site to substrate entry. The structure reveals how zinc embedded in an integral membrane protein can catalyze peptide cleavage.

Regulated intramembrane proteolysis (RIP) is a conserved signaling mechanism from bacteria to humans (*1*–*8*). An essential step of RIP is the site-specific cleavage of a transmembrane signaling protein by a specific membrane-embedded protease within the lipid bilayer. These intramembrane proteases are classified into four families: the metalloprotease site-2 protease (S2P), serine protease rhomboid, and aspartyl proteases presenilin and signal-peptide peptidase (*2*–*4*).

RIP signaling is exemplified by cleavage of the membrane-bound transcriptional factor sterol regulatory element–binding protein (SREBP) by S2P in mammals (*9*–*11*). In response to low levels of cellular cholesterol, SREBP is translocated from the endoplasmic reticulum (ER) to the Golgi, where it is cleaved at a site that is three amino acids into the transmembrane segment on the cytosolic side (*12*). This cleavage follows a prior cleavage in the lumen of the Golgi by the site-1 protease (S1P) (*13*). Consequently, the DNA binding and transactivation domain of SREBP is released from the Golgi membrane and translocated into the nucleus, where it activates transcription of genes that control biosynthesis and uptake of cholesterol and fatty acids. The mammalian S2P is also responsible for cleavage activation of the transcription factor ATF6 (*14*), which plays a central role in the ER stress signaling. In response to periplasmic stress, the *Escherichia coli* S2P homolog YaeL (also known as RseP) cleaves a transmembrane protein RseA, following an initial cleavage event mediated by the periplasmic serine protease DegS (*15*). The *Bacillus subtilis* S2P homolog sporulation protein SpoIVFB removes the prosequence from Pro-σ^K, which allows the resulting transcription factor to activate genes that are required for sporulation (*16*, *17*).

The S2P family proteases contain a consensus HExxH sequence, in which the two histidine residues are thought to coordinate a zinc atom together with a conserved aspartate residue (*1*, *18*–*20*). During catalysis, the conserved glutamate residue is thought to activate a zinc-bound water molecule to initiate nucleophilic attack on the scissile peptide bond. These putative catalytic residues are predicted to be located below the lipid membrane surface. In this case, because proteolysis requires water molecules, how do hydrophilic water molecules enter the active site of S2P? More important, how do transmembrane substrate proteins gain access to the active site? Are there common principles that govern different families of intramembrane proteases? These important questions remain unanswered.

Recent structural investigations on the rhomboid serine proteases revealed tantalizing clues about how an intramembrane protease might function (*21*–*24*). However, because S2P shares no apparent sequence homology with rhomboid, information derived from rhomboid proteases cannot be directly applied to the understanding of S2P. In this study, we report the crystal structure of an S2P homolog from the archaebacterial species *Methanocaldococcus jannaschii*.

Crystallization of mjS2P. We cloned 40 S2P homologs from 31 bacterial and archaebacterial species and examined their expression in *E. coli* (*25*). On the basis of solution behavior, we focused on the S2P homolog from *M. jannaschii* and generated crystals of its transmembrane core domain (residues 1 to 224), hereafter referred to as mjS2P.

We reconstituted a proteolysis assay in vitro, in which mjS2P cleaved an artificial protein substrate CED-9 (*26*) in detergent micelles (Fig. 1A, left). Next, quantitative element analysis revealed that zinc is bound to mjS2P in an approximately 1:1 molar ratio. Finally, the metalloprotease-specific inhibitor 1,10-phenanthroline specifically inhibited substrate cleavage in a concentration-dependent manner (Fig. 1A, right). These analyses validated the use of the transmembrane core domain for crystallographic studies.

The structure of mjS2P was determined by multi-wavelength anomalous dispersion (MAD), with the use of seleno-methionine–labeled protein, and refined to 3.3 Å resolution (table S1 and fig. S1). Details of crystallization and structural determination are given in the supporting online materials (*25*).

Overall structure of mjS2P. There are two molecules of mjS2P in an asymmetric unit, designated A and B, which associate with each

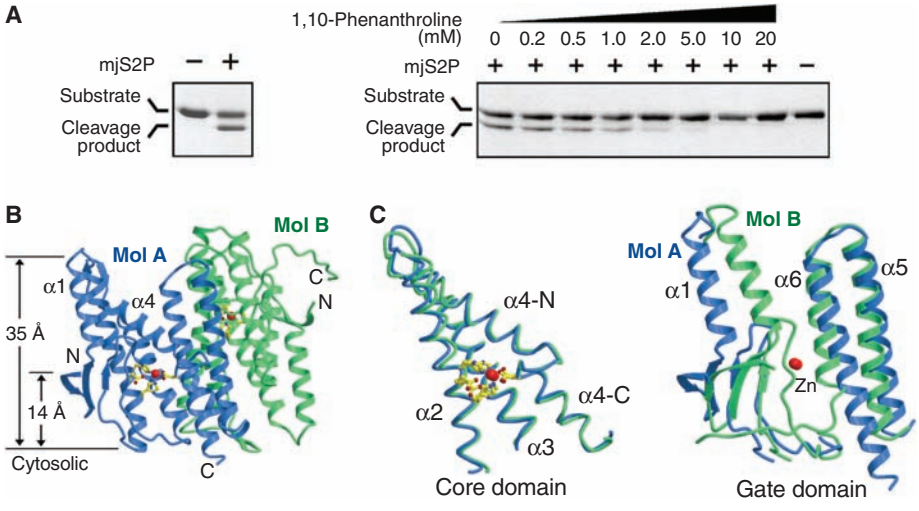


Fig. 1. Structure of the transmembrane core domain of an S2P homolog from *M. jannaschii* (mjS2P). (**A**) The transmembrane core domain of mjS2P is catalytically active. The membrane-associated protein CED-9 (*26*) was used as an artificial protein substrate for mjS2P. The proteolytic activity is inhibited by 1,10-phenanthroline, an inhibitor specific for metalloproteases (right). (**B**) Overall structure of mjS2P in one asymmetric unit. Two molecules of mjS2P, named A (blue) and B (green), associate with each other to form a pseudo-dimer in the crystals. The catalytic zinc atom is highlighted in red. The zinc-binding and catalytic residues are colored yellow. (**C**) The core domain and gate domain of mjS2P. Overlay of the two molecules of mjS2P reveals a shared core domain (left) and a diverging gate domain (right). With the exception of Figs. 3A, 4A, and 4B, all structural figures were made using MOLSCRIPT (*30*).

[1]Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Princeton, NJ 08544, USA. [2]Department of Chemistry, University of British Columbia, Vancouver, BC V6T 1Z3, Canada.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: ygshi@princeton.edu

other to form an antiparallel, pseudo-dimer in the crystals (Fig. 1B). The orientation of each molecule relative to the lipid membrane was assigned on the basis of the locations of charged amino acids between adjacent transmembrane segments. Each molecule contains seven α helices and five short β strands (fig. S2). At the N terminus, two antiparallel β strands (β1 and β2) combine with a third strand β3, located between helices α2 and α3, to form a membrane-embedded β sheet. Strands β4 and β5 form a β hairpin between helices α3 and α4-N.

The secondary structural elements of mjS2P are arranged into a six-transmembrane–segment (TM1–6) topology that does not resemble any known metalloprotease (fig. S2). TM1 consists of strand β2 and helix α1, whereas TM4 contains two separate α helices, α4-N and α4-C, which are connected by a nine–amino acid bulge. TM2, TM3, TM5, and TM6 correspond to helices α2, α3, α5, and α6, respectively. As previously hypothesized, the catalytic zinc atom is coordinated by three amino acids that are invariant among all S2P family proteins: His^{54} and His^{58} from helix α2 and Asp^{148} from the N-terminal end of helix α4-C in mjS2P. The zinc atom is located ~14 Å below the lipid membrane surface from the cytosolic side.

The overall structure of the two mjS2P molecules in one asymmetric unit is similar, with a root-mean-square deviation (RMSD) of 1.8 Å over 150 aligned Cα atoms out of a total of 217 amino acids. In particular, TM2, TM3, and TM4 stack closely against each other and share nearly identical conformations in these two molecules (Fig. 1C, left). However, there are also significant conformational differences. Compared with molecule B, molecule A adopts a relatively open conformation, in which TM1 and TM6 move away from each other by approximately 10 to 12 Å (Fig. 1C, right). This difference results in the exposure of the active site in molecule A, but not in B. From these observations, molecules A and B are proposed to exist in open and closed conformations, respectively.

The conserved conformation of TM2 to 4 is likely essential to the formation of the active site, because His^{54} and His^{58} reside in TM2 and Asp^{148} is located in TM4. TM3 plays a structural role by interacting with and supporting the conformations of TM2 and TM4. In support of this analysis, TM2, TM3, and TM4 have been predicted to be common to all S2P family members (*19*), which contain conserved sequences in these TMs (fig. S2). Thus, we term TM2, TM3, and TM4 the "core domain" of mjS2P. In contrast, TM1, TM5, and TM6 exhibit different conformations in the two mjS2P molecules and may represent two distinct states in the regulation of substrate entry. We thus term TM1, TM5, and TM6 the "gate domain" of mjS2P.

The active site. In molecule B, the distances between zinc and the coordinating atoms of His^{54}, His^{58}, and Asp^{148} are 2.3, 2.2, and 2.1 Å, respectively (Fig. 2A). In molecule A, these distances become 2.3, 2.3, and 2.2 Å, respectively. Coordination of the zinc atom is roughly tetrahedral, with the fourth water ligand unassigned owing to the moderate resolution. In both molecules, His^{58} is hydrogen-bonded to the carbonyl oxygen atom of residue 94. The closest carboxylate oxygen atom of Glu^{55} is 3.3 and 3.4 Å away from the zinc atom in molecules A and B, respectively. These distances are consistent with activation by Glu^{55} of the zinc-bound water molecule during catalysis. It is noteworthy that Asn^{140}, another invariant residue among S2P family proteases, is located above the active site in the open space. This location suggests a critical role for Asn^{140} in catalysis, perhaps in binding to substrate and/or helping the formation of the oxyanion hole. In support of this notion, comparison of the active-site conformation between mjS2P and thermolysin, a HExxH-containing metalloprotease, revealed that Asn^{140} is located in approximately the same position as Arg^{203} in thermolysin (*27*) (Fig. 2B), which binds to the carbonyl oxygen of the scissile peptide bond and neutralizes the negative charge during catalysis (*28*).

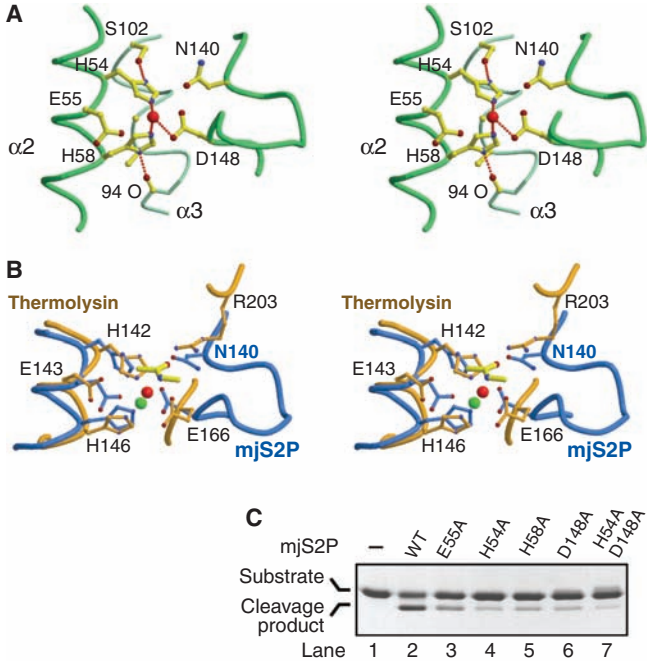


Fig. 2. Conformation of the active site. (**A**) A stereo view of the active-site conformation in molecule B. The catalytic zinc atom is coordinated by His^{54} and His^{58} on helix α2 and Asp^{148} at the N-terminal end of helix α4-C. Glu^{55} likely facilitates the activation of a water molecule during catalysis. The conserved Asn^{140} may contribute to the formation of the oxyanion hole. Helix α3 does not directly participate in catalysis but interacts with helices α2 and α4 to form the conserved core domain. (**B**) A stereo comparison of the active-site conformation between mjS2P (blue) and a HExxH-containing metalloprotease thermolysin [gold, PDB code 2TLX (*27*)]. Asn^{140} of mjS2P is located in a similar position as Arg^{203} in thermolysin, which is part of the oxyanion hole. (**C**) Mutation of the active-site residues compromised proteolytic activity. Shown here is an SDS-PAGE gel visualized by Coomassie staining. CED-9 was used as an artificial substrate in these assays.

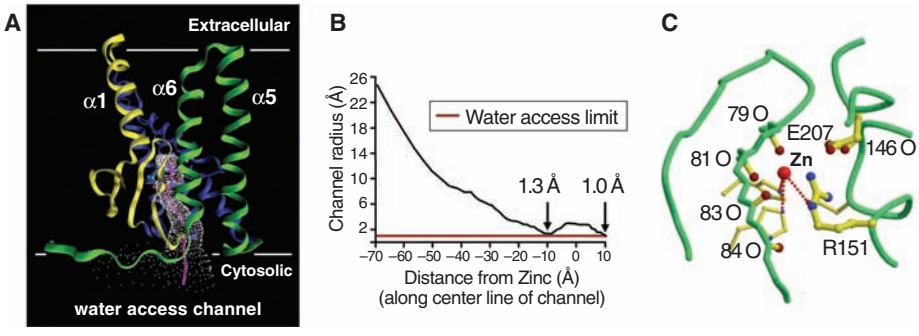


Fig. 3. Access of water molecules to the active site of mjS2P. (**A**) The van der Waals surface in molecule B reveals a channel that leads to the active site from the cytosolic side. The calculation was performed with the program HOLE (*31*) and the image was generated using VMD (*32*). The center line of the channel is colored magenta. (**B**) The channel is large enough to allow passage of water molecules. Distance from the zinc atom along the center line of the channel is plotted against the minimal radius at each point. The red line indicates the minimal radius required for passage of water molecules. (**C**) The channel is lined with polar groups that may help facilitate water entry. Shown here is a view of the polar groups in the channel approximately along the center line. Five carbonyl oxygen atoms, Arg^{151}, and Glu^{207} are positioned along the inside of the channel.

To assess the contribution of the catalytic residues, we generated missense mutations in mjS2P. The mutant proteins H54A (in which Ala replaces the His at residue 54), E55A, H58A, D148A, and H54A/D148A (*20*) exhibited markedly compromised protease activity compared with the wild-type (WT) mjS2P (Fig. 2C). Quantitative element analysis by inductively coupled plasma-emission spectrometry revealed that, compared with the WT protein, the zinc content was 8.2, 44.2, 5.1, and 15.1% for the mutant proteins H54A, E55A, H58A, and D148A, respectively.

An important structural feature of mjS2P is that all active-site residues are contained within TM2 and TM4, with TM3 stabilizing the active-site conformation from the opposite side of where potential substrate proteins are cleaved (Fig. 2A). This arrangement immediately rules out the possibility that conserved residues in TM3 may directly participate in catalysis. Consistent with this analysis, Ala^{97} and Gly^{98} in the conserved AGxxxN/S/G sequence of TM3 (*19*) appear to play a structural role in mjS2P. For example, Gly^{98} stacks closely against His^{54} and His^{58}, and its substitution for any other amino acid is predicted to perturb the conformation of the active site because of steric hindrance from the side chain.

Access to water molecules. How do water molecules gain access to the active site of mjS2P? Analysis of the van der Waals surface in molecule B reveals a channel that originates from the zinc atom to the cytosolic side (Fig. 3A). The narrowest point in this channel measures ~2.6 Å in diameter (Fig. 3B), which is large enough to allow passage of a water molecule. Notably, the inner surface of the channel contains a number of polar groups and charged amino acids that may facilitate water entry. Five carbonyl oxygen atoms from residues 79, 81, 83, 84, and 146 point into the channel, poised to coordinate water molecules (Fig. 3C). In addition, the side chains of two charged amino acids Arg^{151} and Glu^{207} also point into the channel. Thus, water molecules appear to have constant access to the active site of mjS2P in the closed conformation.

Mechanism of substrate entry. Molecules A and B exhibit markedly different conformations in TM1. In the closed molecule B, exclusively hydrophobic amino acids from TM1 and TM2 interdigitate to form an extensive network of van der Waals interactions (fig. S3A). Compared with molecule B, TM1(helix α1 and strand β2) pivots outward around helix α2 by ~35° to reach its position in molecule A (fig. S3B). This movement results in major alteration of the hydrophobic interface between TM1 and TM2. TM6 and TM5 also have different conformations in molecules A and B, although such differences are small compared with those in TM1. Unlike TM1, conformational changes in TM6 and TM5 do not result in the repacking of hydrophobic interfaces, because helix α4-C also slightly adjusts its position to maintain the same packing interactions with α6 and α5.

The two contrasting conformations of mjS2P molecules provide a plausible explanation for the question of how substrate proteins gain access to the active site. In the closed molecule B, the active site is inaccessible to transmembrane protein substrate (Fig. 4A). In the open molecule A, there is a deep groove that is roughly parallel to the transmembrane helices (Fig. 4B). This groove traverses through the entire molecule and exposes the active site to potential substrate peptide. These observations suggest a lateral gating mechanism, in which TM1 and TM6-TM5, as two sides of the gate, move away from each other to allow substrate entry and catalysis (Fig. 4C).

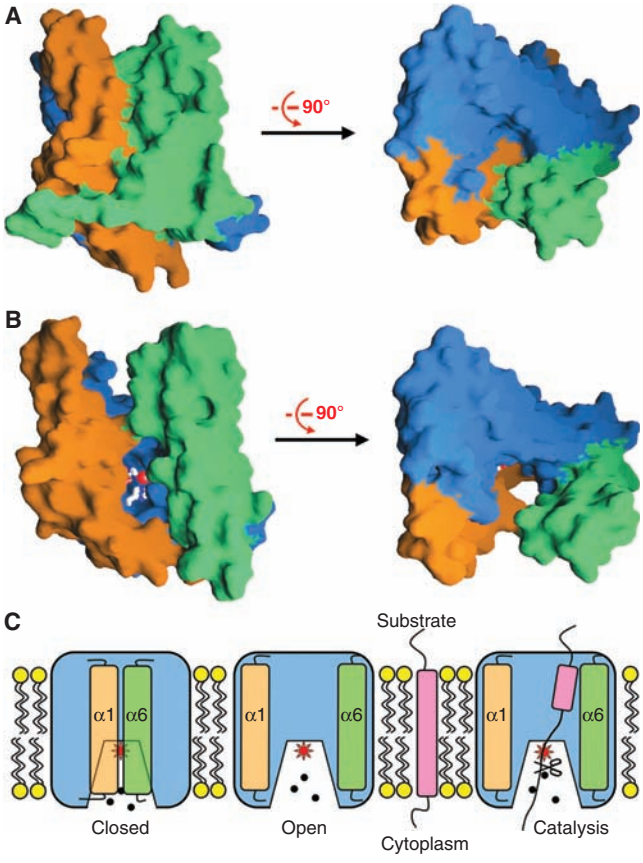


Fig. 4. Mechanism of substrate gating in mjS2P. (**A**) Surface representation of the closed state of mjS2P in two perpendicular views. Note the closure of the active site. (**B**) Surface representation of the open state of mjS2P in two perpendicular views. In this conformation, an extended polypeptide can be readily fitted into the cleft between the two gating helices (α1 and α6). (A) and (B) were prepared using GRASP (*33*). (**C**) A proposed general model for the S2P family of intramembrane proteases. In this model, substrate entry to the active site is gated by two transmembrane segments, TM1 and TM6-TM5.

Discussion. How is a substrate protein recognized by S2P? Although the current study does not provide a direct answer to this question, it reveals some tantalizing clues. For example, a number of buried amino acids in the closed conformation (molecule B) become exposed in the open conformation (molecule A), thus creating novel surface features that might be involved in binding to substrate proteins. Compared with the bottom half of the putative substrate-binding groove, the top half is much wider (Fig. 4B) and could accommodate an intact α-helix. We speculate that the funnel-shaped groove in the open conformation may play an active role in unwinding the transmembrane helix of the substrate protein.

Previous studies show that substrate cleavage of the putative transmembrane helix occurs in various positions. These observations appear to contrast with the prediction that the catalytic zinc atom and the active-site residues are likely located in approximately the same depth into the membrane for S2P proteases. This discrepancy can be reconciled by the hypothesis that specific cleavage of a substrate protein is determined by specific recognition.

Structure determination of mjS2P allows comparison with the rhomboid proteases, which represent the only other structurally characterized intramembrane protease (*21*–*24*). Despite the superficial similarity of 6TMs, the structures of mjS2P and the rhomboid serine protease GlpG exhibit different topologies and share no apparent features. Nonetheless, water molecules appear to gain access to the active sites of rhomboid and mjS2P via a similar mechanism. GlpG from *E. coli* contains a water-filled cavity that converges on the active-site residue Ser^{201} and opens to the extracellular side (*21*–*23*). In contrast, mjS2P has a polar channel that allows water entry to the catalytic zinc atom in the closed conformation. Both classes of intramembrane proteases also appear to share a common feature in mechanisms of substrate entry—gating by transmembrane helices, although significant differences exist. In GlpG, bending of the C-terminal half of TM5 was proposed to open the gate for substrate entry (*22*). This hypothesis is consistent with other structural observations (*21*, *23*, *24*) and biochemical characterization (*29*). In mjS2P, the rotation and translocation of TM1 and the translocation of TM6-TM5 may allow substrate entry.

The full-length mjS2P protein contains 339 amino acids; but only the transmembrane core

domain (residues 1 to 224) was used for crystallization. Could the structure be altered or disrupted by not including the C-terminal sequences? Two lines of evidence argue against such a possibility. First, the C-terminal sequences of mjS2P were frequently lost during bacterial expression, which suggests that they are unlikely to be part of the structural core domain. Indeed, the transmembrane core domain of mjS2P was identified by using limited proteolysis. Second, the transmembrane core domain retained full proteolytic activity compared with the full-length mjS2P (fig. S4).

Because mjS2P was crystallized in the presence of detergents, it is possible that the observed conformations are, in part, induced by interactions with the detergent molecules or crystal-packing interactions. Although we could not rule out this possibility, the observed active-site geometry and the conserved core domain structure in the two molecules do not suggest anything had gone awry. Nonetheless, further experiments will determine whether the proposed closed and open conformations of mjS2P in the crystals are physiologically relevant. The large conformational differences of the gate domain in the two molecules are consistent with the fact that TM1 and TM6-TM5 make up highly divergent sequences among members of the S2P family. This feature, also observed in TM5 of GlpG, would better allow specificity for recognition and cleavage of substrate proteins.

References and Notes

1. M. S. Brown, J. Ye, R. B. Rawson, J. L. Goldstein, *Cell* **100**, 391 (2000).
2. S. Urban, M. Freeman, *Curr. Opin. Genet. Dev.* **12**, 512 (2002).
3. M. S. Wolfe, R. Kopan, *Science* **305**, 1119 (2004).
4. J. O. Ebinu, B. A. Yankner, *Neuron* **34**, 499 (2002).
5. M. Ehrmann, T. Clausen, *Annu. Rev. Genet.* **38**, 709 (2004).
6. R. B. Rawson, *Essays Biochem.* **38**, 155 (2002).
7. H. Makinoshima, M. S. Glickman, *Microbes Infect.* **8**, 1882 (2006).
8. S. Urban, *Genes Dev.* **20**, 3054 (2006).
9. J. Sakai *et al.*, *Cell* **85**, 1037 (1996).
10. M. S. Brown, J. L. Goldstein, *Cell* **89**, 331 (1997).
11. R. B. Rawson, *Mol. Cell* **1**, 47 (1997).
12. E. A. Duncan, U. P. Dave, J. Sakai, J. L. Goldstein, M. S. Brown, *J. Biol. Chem.* **273**, 17801 (1998).
13. J. Sakai *et al.*, *Mol. Cell* **2**, 505 (1998).
14. J. Ye *et al.*, *Mol. Cell* **6**, 1355 (2000).
15. B. M. Alba, J. A. Leeds, C. Onufryk, C. Z. Lu, C. A. Gross, *Genes Dev.* **16**, 2156 (2002).
16. D. Z. Rudner, P. Fawcett, R. Losick, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14765 (1999).
17. Y. T. Yu, L. Kroos, *J. Bacteriol.* **182**, 3305 (2000).
18. A. P. Lewis, P. J. Thomas, *Protein Sci.* **8**, 439 (1999).
19. L. N. Kinch, K. Ginalski, N. V. Grishin, *Protein Sci.* **15**, 84 (2006).
20. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, any amino acid; and Y, Tyr.
21. Y. Wang, Y. Zhang, Y. Ha, *Nature* **444**, 179 (2006).
22. Z. Wu *et al.*, *Nat. Struct. Mol. Biol.* **13**, 1084 (2006).
23. A. Ben-Shem, D. Fass, E. Bibi, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 462 (2007).
24. M. J. Lemieux, S. J. Fischer, M. M. Cherney, K. S. Bateman, M. N. James, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 750 (2007).
25. Materials and methods are available as supporting material on *Science* online.
26. M. O. Hengartner, H. R. Horvitz, *Cell* **76**, 665 (1994).
27. A. C. English, S. H. Done, C. R. Groom, R. E. Hubbard, *Proteins* **37**, 628 (1999).
28. B. M. Matthews, *Acc. Chem. Res.* **21**, 333 (1988).
29. R. P. Baker, K. Young, L. Feng, Y. Shi, S. Urban, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8257 (2007).
30. P. J. Kraulis, *J. Appl. Cryst.* **24**, 946 (1991).
31. O. S. Smart, J. M. Goodfellow, B. A. Wallace, *Biophys. J.* **65**, 2455 (1993).
32. W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **14**, 33 (1996).
33. A. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).

34. We thank A. Saxena at Brookhaven National Laboratory, National Synchrotron Light Source beamlines for help. The atomic coordinates of the transmembrane core domain of mjS2P have been deposited in the Protein Data Bank with the accession code 3B4R.

20 September 2007; accepted 29 October 2007
10.1126/science.1150755

REPORTS

A Cosmic Microwave Background Feature Consistent with a Cosmic Texture

M. Cruz,[1,2]* N. Turok,[3] P. Vielva,[1] E. Martínez-González,[1] M. Hobson[4]

The Cosmic Microwave Background provides our most ancient image of the universe and our best tool for studying its early evolution. Theories of high-energy physics predict the formation of various types of topological defects in the very early universe, including cosmic texture, which would generate hot and cold spots in the Cosmic Microwave Background. We show through a Bayesian statistical analysis that the most prominent 5°-radius cold spot observed in all-sky images, which is otherwise hard to explain, is compatible with having being caused by a texture. From this model, we constrain the fundamental symmetry-breaking energy scale to be $\phi_0 \approx 8.7 \times 10^{15}$ gigaelectron volts. If confirmed, this detection of a cosmic defect will probe physics at energies exceeding any conceivable terrestrial experiment.

The Cosmic Microwave Background (CMB) radiation was emitted from the hot plasma of the early universe roughly 14 billion years ago. All-sky, multifrequency maps of the CMB sky made by the Wilkinson Microwave Anisotropy Probe (WMAP) (*1*, *2*) reveal Gaussian temperature anisotropies of the form expected in standard cosmological scenarios, tracing density variations of one part in a hundred thousand in the primordial cosmos (*3*). However, several apparent anomalies in the expected Gaussian, isotropic statistical distribution have also been found (*4*–*10*). One of the most striking is a large cold spot centered on Galactic coordinates $b = -57°$, $l = 209°$, with a radius of $\approx 5°$ (*10*–*13*). It was detected using the Spherical Mexican Hat Wavelet (SMHW), an optimal tool for enhancing such features, and has a flat frequency spectrum, inconsistent with either Galactic foregrounds or the Sunyaev-Zel'dovich effect (*13*). A conservative estimate of the probability of finding such a feature in Gaussian simulations, taking the effect of a posteriori selection into account, is only 1.85%. Several radical explanations, such as huge voids or an anisotropic cosmology, have already been proposed: Many have been ruled out by other cosmological observations (*14*–*17*).

Here, we consider the possibility that the spot was caused by a cosmic texture (*18*), a type of cosmic defect predicting spots in the CMB (*19*). Cosmic defects are hypothetical remnants of symmetry-breaking phase transitions in the early universe, predicted by certain unified theories of elementary particle physics. According to these theories, the different species of elementary particle are indistinguishable in the hot early universe. As the universe cools, the symmetry between them breaks, in a phase transition analogous to the freezing of water. Just as misalignments in the crystalline structure of ice lead to defects, misalignments in the symmetry breaking in unified theories lead to the formation of cosmic defects (*20*, *21*). Breaking a discrete symmetry produces domain walls and breaking a circle [or $U(1)$] symmetry produces cosmic strings. Textures form when a

[1]Instituto de Física de Cantabria, Consejo Superior de Investigaciones Cientifícas Univ. de Cantabria, Avenida los Castros, 39005-Santander, Spain. [2]Departamento de Física Moderna, Universidad de Cantabria, Avenida los Castros, 39005-Santander, Spain. [3]Department of Applied Mathematics and Theoretical Physics, Center for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK. [4]Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK.

*To whom correspondence should be addressed. E-mail: cruz@ifca.unican.es

simple Lie group [like the special unitary group, $SU(2)$] is broken. They consist of localized, twisted configurations of fields which collapse and unwind on progressively larger scales, a scaling process continuing into the late universe. Each unwinding texture creates a concentration of stress-energy and a time-varying gravitational potential. CMB photons passing through such a region receive a red- or blueshift, creating a cold or hot spot (*19*) with a magnitude set by the symmetry-breaking energy scale ϕ_0. We have used high-resolution texture simulations and Bayesian statistical analysis to investigate whether the observed spot is consistent with a texture. We conclude that it is, and we propose further tests. If confirmed, the detection will provide a window onto physics at extremely high energies.

Texture unwinding events may be approximated by a spherically symmetric scaling solution (*21*) on a comoving radius $r < \kappa\tau$, where κ is a fraction of unity and τ is the conformal time when the texture unwinds. Such events lead to hot and cold spots, with a fractional temperature distortion

$$\frac{\Delta T}{T}(\theta) = \pm\varepsilon \frac{1}{\sqrt{1+4\left(\frac{\theta}{\theta_C}\right)^2}} \qquad (1)$$

where θ is the angle from the center, the amplitude is set by $\varepsilon = 8\,\pi^2\,G\,\phi_0^2$ and the scale parameter $\theta_C \equiv 2\,\sqrt{2}\,\kappa(1+z)/\,(E(z)\int_0^z [(d\,\overline{z})/E(\overline{z})])$ and $E(z) = \sqrt{\{\Omega_M\,(1+z)^3 + \Omega_\Lambda\}}$, with Ω_M and Ω_Λ the present-day matter and dark-energy density parameters and z the redshift of the unwinding texture. Because the scaling profile is not valid at large comoving radii r, we truncate Eq. 1 beyond its half-maximum by matching its value and derivative to a Gaussian function. A scale-invariant distribution of spots on the sky is predicted (*22*), with the number of spots of scale θ_C or above, $N_{spot}(>\theta_C) = 4\,\pi\nu\kappa^3\,/(3\theta_C^2)$. Here, ν parameterizes the comoving number density n of unwinding textures: $dn/d\tau = \nu\tau^{-4}$. High-resolution simulations of $SU(2)$ textures yield $\kappa \approx 0.1$ and $\nu \approx 2$ (*23*). The corresponding fraction of the sky covered by textures with θ_C larger than 1° is $f_S \approx 0.017$.

We perform a Bayesian analysis of the probability ratio ρ of two hypotheses given the data [see, e.g., (*24*)]. The null hypothesis H_0 describes the data D as a statistically homogeneous and isotropic Gaussian random field (CMB) plus instrument noise. The alternative hypothesis H_1 describes the data as CMB plus noise and an additional template, T, given by a cosmic texture with parameters ε and θ_C. The evidence is the average of the likelihood L with respect to the prior, Π: $E_i = Pr(\mathrm{D}|H_i) = \int L_i(\Theta_i|H_i)\Pi(\Theta_i)d\Theta_i$, where the Θ_i are the parameters in hypothesis H_i. This formula naturally incorporates an Occam factor favoring the hypothesis with fewer parameters. Our template fitting is performed in a circular area of 20° radius centered on $b = -57°$, $l = 209°$, in the three-year foreground-cleaned WMAP combined-frequency map (WCM) (*1*) at ≈ 1° resolution (*22*).

The posterior probability ratio $\rho = Pr(H_1|\mathrm{D})/Pr(H_0|\mathrm{D}) = E_1\,Pr(H_1)/[E_0\,Pr(H_0)]$ can be used to decide between the hypotheses. The alternative hypothesis is favored when $\rho > 1$ and rejected otherwise. The a priori probability ratio for the two models, $Pr(H_1)/Pr(H_0)$, is usually set to unity, but because we center the template at an a posteriori selected pixel, we set $Pr(H_1)/Pr(H_0)$ to the fraction of sky covered by textures. To compute E_1/E_0, we need the likelihood and normalized priors. The likelihood function is just $L \propto \exp(-\chi^2/2)$, where $\chi^2 = (\mathrm{D}-\mathrm{T})'\,\mathrm{N}^{-1}\,(\mathrm{D}-\mathrm{T})$ and N is the generalized noise matrix, including CMB and instrument noise (*22*).

As a conservative prior on ε, we choose $0 \le \varepsilon \le 10^{-4}$, the latter value being the upper limit imposed by the large-scale Cosmic Background Explorer (COBE) satellite measurements (*25*–*27*). The prior for θ_C is obtained by normalizing the number of spots distribution, $dN_{spot}/d\theta_C \propto \theta_C^{-3}$, to unity between θ_{min} and θ_{max}. Photon diffusion would smear out textures smaller than a degree or so, hence we set $\theta_{min} = 1°$. At large scales, textures are unlikely because the sky is finite: We set $\theta_{max} = 15°$.

We find the probability ratio $\rho \approx 2.5$, favoring the texture plus Gaussian CMB over the Gaussian-only model. The data, the best-fit template, and the difference between the two are presented in Fig. 1. The best-fit amplitude and scale are $\varepsilon = 7.7\times10^{-5}$ and $\theta_C = 5.1°$. Marginalizing the posterior (i.e, likelihood times prior; see Fig. 2), we find $\theta_C = 4.9^{\circ\,+2.8}_{\;-2.4}$ and $\varepsilon = 7.3^{+2.5}_{-3.6}\times10^{-5}$ at 95% confidence. The value of ε inferred in this way from a single extreme event is biased by the detection of signals with high noise, that is, large Gaussian fluctuations. To check this, we generated 500 all-sky, Gaussian CMB simulations (*10*, *22*) and added one cold texture spot to each, with amplitude $\varepsilon = 4 \times 10^{-5}$ below the upper limit, 5×10^{-5} inferred from the observed CMB anisotropy spectrum (*28*). We perform the same template fit we applied to the data on each cold spot and then select the spots with high posterior probability ratios, $\rho > 1$. The mean amplitude obtained from these spots is $\varepsilon \approx 7.9 \times 10^{-5}$; hence, there is considerable overestimation. Moreover, a more realistic model of textures would predict some dispersion in the spot strength, with stronger spots caused by asymmetric, multiple, or moving textures. Again, estimating ε from the strongest texture spot would lead to a biased value.

As a complementary test for the a posteriori selection of the template center, we should also study whether prominent Gaussian CMB spots produce such high values of ρ. Following the same procedure for 10,000 Gaussian simulations, we select the most prominent spot, finding that these spots show typical values of $\rho \approx 0.14 < 1$, with only ~ 5.8% of the simulations showing spots with $\rho > 2.5$. Because the kurtosis of the data shows a more appreciable departure from Gaussianity, the percentage might further decrease if spots of all sizes were taken into account.

To analyze further the CMB signal from textures, we generate 10,000 texture plus Gaussian

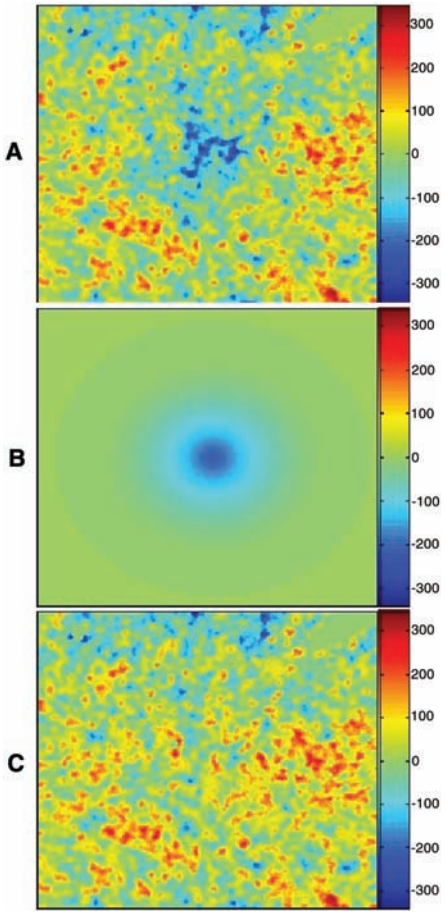


Fig. 1. (**A**) Azimuthal projection of a 43° × 43° patch of the WCM, centered on ($b = -57°$, $l = 209°$). (**B**) Best-fit texture template. (**C**) WCM subtracting the texture template. The temperature units shown in the color bars are μK and the pixel is 13.7 arcmin. The y axis points to the Galactic north pole. (The template is available on request).

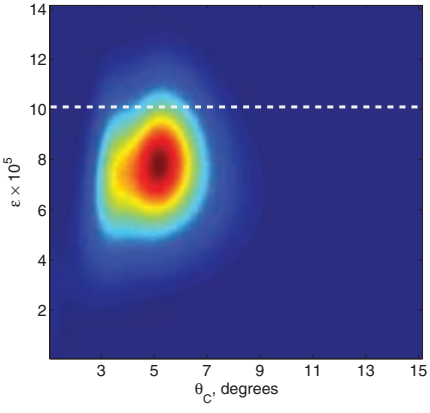


Fig. 2. Posterior (likelihood times prior) as a function of amplitude ε and scale θ_C for the texture template fit. The prior limit on the amplitude is marked by a dashed white line.

CMB and noise simulations and repeat the analyses performed on Gaussian simulations with no textures. Considering that we observe one 5.1° texture in about 40% of the sky of the WCM—the region unaffected by the Galaxy—we predict ~68 hot and cold spots above θ_{min} = 1°. We generate 68 spots per simulation with $N_{spot}(> \theta_C) \propto \theta_C^{-2}$ and θ_{min} = 1°, assigning a random sign and position to each spot in the sky. The amplitude is set to $\varepsilon = 4 \times 10^{-5}$ following the discussion above. We then repeat the previously performed multiscale analysis of skewness and kurtosis (*10*, *11*). Because there are on average the same number of hot and cold spots, the skewness is little affected. On the contrary, however, the kurtosis is increased so the anomalously high kurtosis of the data at scales around 5° is actually compatible with the Gaussian CMB plus textures interpretation (Fig. 3).

From the results of the texture simulations, the predicted number of spots $N_{spot}(> \theta_C) \approx 1.1$ for $\theta_C \approx 5.1°$, consistent with the single observed spot. Hence, we find that the abundance, shape, size and amplitude of the spot are consistent with the texture interpretation. The symmetry-breaking scale corresponding to the inferred amplitude is $\phi_0 \approx 8.7 \times 10^{15}$ *GeV*. From the relations given below Eq. 1, the observed texture unwound at $z \sim 6$, after the reionization of the intergalactic medium and potentially within reach of very deep Galaxy or quasar surveys.

Further observations could test the texture hypothesis. First, if the spot is due to a texture, it was caused by a time-dependent gravitational potential. There would be no associated CMB polarization. However, if the spot is a rare statistical fluctuation in the primordial density, a correlated polarization signal, namely a preference for a radial pattern of CMB polarization around it, is expected. On these scales, for adiabatic perturbations with standard recombination, almost half the polarization signal is correlated with the temperature anisotropy (*29*).

Second, there should be many smaller texture spots, with the scale-invariant distribution described above. These would be masked or confused by the background Gaussian signal where it has maximal power, at $\theta \sim 1°$ to 2°. Nevertheless, each spot would deviate from the expected polarization-temperature correlation, and a combined all-sky measurement might show a considerable difference from the standard prediction.

Finally, a texture at $z \sim 6$ would gravitationally lens objects behind it with a lensing angle on the order of ε radians. In particular, it would lens the second-order CMB anisotropies (the Vishniac effect), which peak at these angular scales.

Although certainly radical, the texture hypothesis is the most plausible explanation yet proposed for the spot. Our analysis shows it to be favored over a purely Gaussian CMB. Alternate explanations, such as voids with radius ~300 Mpc, are far more radical and seem inconsistent both with standard cosmology and with Galaxy survey observations (*30*).

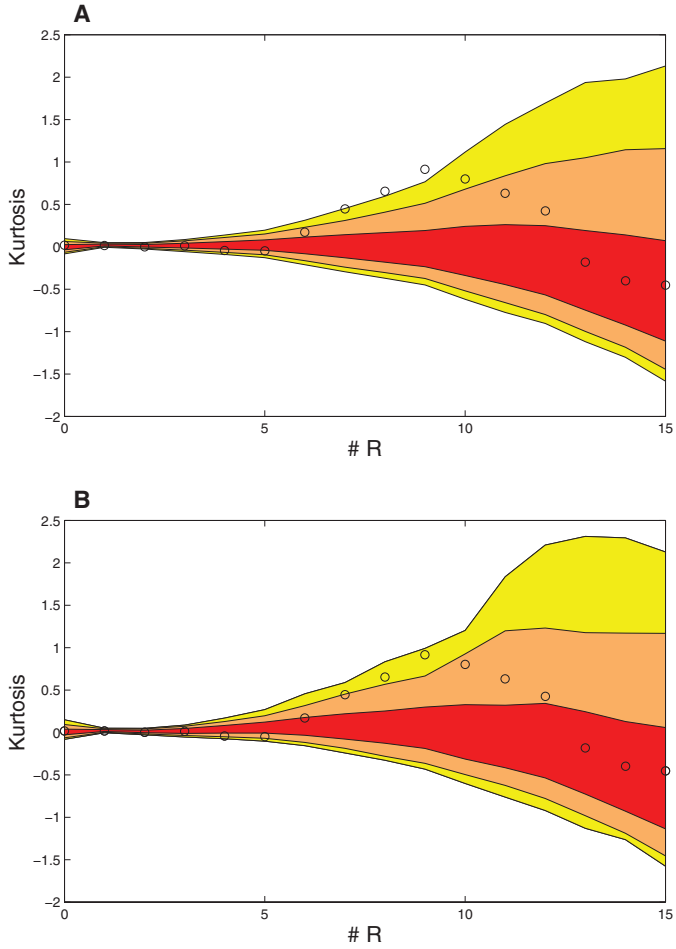


Fig. 3. Kurtosis of the WCM convolved with the SMHW at 15 scales. The bands represent the 68% (red), 95% (orange), and 99% (yellow) acceptance intervals given by 10,000 simulations of Gaussian CMB (**A**) and Gaussian CMB plus textures (**B**). The WCM data (circles) show deviation from the expected values compared with Gaussian simulations but are fully consistent with the Gaussian CMB plus textures interpretation at all scales. # *R* stands for scale number. The 15 scales are R_1 = 13.7, R_2 = 25, R_3 = 50, R_4 = 75, R_5 = 100, R_6 = 150, R_7 = 200, R_8 = 250, R_9 = 300, R_{10} = 400, R_{11} = 500, R_{12} = 600, R_{13} = 750, R_{14} = 900 and R_{15} = 1050 arcmin.

References and Notes

1. C. L. Bennett *et al.*, *Astrophys. J. Suppl.* **148**, 1 (2003).
2. G. Hinshaw *et al.*, *Astrophys. J. Suppl.* **170**, 288 (2007).
3. D. Spergel *et al.*, *Astrophys. J. Suppl.* **170**, 377 (2007).
4. A. de Oliveira-Costa, M. Tegmark, *Phys. Rev. D* **74**, 023005 (2006).
5. C. J. Copi, D. Huterer, D. J. Schwarz, G. D. Starkman, *Phys. Rev. D* **75**, 023507 (2007).
6. K. Land, J. Magueijo, *Mon. Not. R. Astron. Soc.* **378**, 153 (2007).
7. J. D. McEwen, M. P. Hobson, A. N. Lasenby, D. J. Mortlock, *Mon. Not. R. Astron. Soc.* **371**, 50 (2006).
8. H. K. Eriksen, A. J. Banday, K. M. Górski, F. K. Hansen, P. B. Lilje, *Astrophys. J.* **660**, 81 (2007).
9. Y. Wiaux, P. Vielva, E. Martínez-González, P. Vandergheynst, *Phys. Rev. Lett.* **96**, 1511303 (2006).
10. M. Cruz, L. Cayón, E. Martínez-González, P. Vielva, J. Jin, *Astrophys. J.* **655**, 11 (2007).
11. P. Vielva, E. Martínez-González, R. B. Barreiro, J. L. Sanz, L. Cayón, *Astrophys. J.* **609**, 22 (2004).
12. M. Cruz, E. Martínez-González, P. Vielva, L. Cayón, *Mon. Not. R. Astron. Soc.* **356**, 29 (2005).
13. M. Cruz, M. Tucci, E. Martínez-González, P. Vielva, *Mon. Not. R. Astron. Soc.* **369**, 57 (2006).
14. K. Tomita, *Phys. Rev. D* **72**, 10 (2005).
15. K. T. Inoue, J. Silk, *Astrophys. J.* **664**, 650 (2007).
16. T. R. Jaffe, S. Hervik, A. J. Banday, K. M. Górski, *Astrophys. J.* **644**, 701 (2006).
17. M. Bridges, J. D. McEwen, A. N. Lasenby, M. P. Hobson, *Mon. Not. R. Astron. Soc.* **377**, 1473 (2007).
18. N. Turok, *Phys. Rev. Lett.* **63**, 2625 (1989).
19. N. Turok, D. N. Spergel, *Phys. Rev. Lett.* **64**, 2736 (1990).
20. T. W. B. Kibble, *J. Phys.* **A9**, 1387 (1976).
21. A. Vilenkin, E. P. S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge Univ. Press, Cambridge, 2000).
22. Materials and methods are available as supporting material on *Science* Online.
23. See www.damtp.cam.ac.uk/cosmos/viz/movies/neil.html.
24. M. P. Hobson, C. McLachlan, *Mon. Not. R. Astron. Soc.* **338**, 765 (2003).
25. U. Pen, D. N. Spergel, N. Turok, *Phys. Rev. D* **49**, 692 (1994).
26. D. P. Bennett, S.-H. Rhie, *Astrophys. J.* **406**, L7 (1993).
27. R. Durrer, M. Kunz, A. Melchiorri, *Phys. Rev. D* **59**, 123005 (1999).
28. N. Bevis, M. Hindmarsh, M. Kunz, *Phys. Rev. D* **70**, 043508 (2004).
29. R. G. Crittenden, D. Coulson, N. Turok, *Phys. Rev. D* **52**, R5402 (1995).
30. F. Hoyle, M. S. Vogeley, *Astrophys. J.* **607**, 751 (2004).

31. M.C. thanks the Ministerio de Educacion Cultura y Ciencia for a predoctoral fellowship. N.T. thanks The Science and Technology Facilities Council (UK) and the Centre for Theoretical Cosmology in Cambridge for support. P.V. thanks an I3P contract from the Consejo Superior de Investigaciones Científicas. M.C., P.V., and E.M.G. thank the Ministerio de Educacion y Ciencia, ref. ESP2004-07067-C03-01. The authors acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Support for LAMBDA is provided by the NASA Office of Space Science.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1148694/DC1
SOM Text
References

2 August 2007; accepted 15 October 2007
Published online 25 October 2007;
10.1126/science.1148694
Include this information when citing this paper.

Modeling the Localized-to-Itinerant Electronic Transition in the Heavy Fermion System $CeIrIn_5$

J. H. Shim,* K. Haule, G. Kotliar

We address the fundamental question of crossover from the localized to the itinerant state of a paradigmatic heavy fermion material: $CeIrIn_5$. The temperature evolution of the one-electron spectra and the optical conductivity are predicted from first-principles calculation. The buildup of coherence in the form of a dispersive many-body feature is followed in detail, and its effects on the conduction electrons of the material are revealed. We find multiple hybridization gaps and link them to the crystal structure of the material. Our theoretical approach explains the multiple peak structures observed in optical experiments and the sensitivity of $CeIrIn_5$ to substitutions of the transition metal element and may provide a microscopic basis for the more phenomenological descriptions currently used to interpret experiments in heavy fermion systems.

Heavy fermion materials have unusual properties arising from the presence of a partially filled shell of *f* orbitals and a very broad band of conduction electrons. At high temperatures, the *f* electrons behave as atomic local moments. As the temperature is reduced, the moments combine with the conduction electrons to form a fluid of very heavy quasiparticles, with masses that are two to three orders of magnitude larger than the mass of the electron (*1*, *2*).

These heavy quasiparticles can undergo superconducting or magnetic transitions at much lower temperatures. Understanding how the itinerant low-energy excitations emerge from the localized moments of the *f* shell is one of the central challenges of condensed-matter physics. It requires the understanding of how the dual, atomic particle–like character and itinerant, wavelike character of the electron manifest themselves in the different physical properties of a material.

$CeIrIn_5$ (*3*) has a layered tetragonal crystal structure (*4*, *5*) (Fig. 1A) in which the layers of $CeIn_3$ (shown as red and gray spheres) are stacked between layers of $IrIn_2$ (shown as yellow and gray spheres). Each Ce atom is surrounded by four In atoms in the same plane and by eight In atoms out of plane.

To describe the electronic structure of these classes of materials, one needs to go beyond the traditional concepts of bands and atomic states and focus on the concept of a spectral function $A(\mathbf{k},\omega)_{LL}$, the function which describes the quantum mechanical probability of removing or adding an electron with angular momentum and atomic character $L = (l, m, a)$, momentum $\mathbf{k}$, and energy ω. It is measured directly in angle-resolved photoemission experiments (ARPES) and inverse photoemission experiments.

Center for Materials Theory, Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA.

*To whom correspondence should be addressed. E-mail: jishim@physics.rutgers.edu

To evaluate the spectral function, we use dynamical mean field theory (DMFT) (*6*) in combination with the local density approximation (LDA+DMFT) (*7*), which can treat the realistic band structure, the atomic multiplet splitting, and Kondo screening on the same footing. The spectral function $A(\mathbf{k},\omega)$ is computed from the corresponding one-electron Green's function $G(\mathbf{k},\omega) = \frac{1}{(\omega+\mu)O_{\mathbf{k}} - H_{\mathbf{k}} - \Sigma(\omega)}$ through the relation $A(\mathbf{k},\omega) = [G^{\dagger}(\mathbf{k},\omega) - G(\mathbf{k},\omega)]/(2\pi i)$. Here, $H_{\mathbf{k}}$ and $O_{\mathbf{k}}$ are the Hamiltonian and overlap matrix obtained by the LDA method (*8*), μ is the chemical potential, i is the imaginary unit, $G^{\dagger}$ is the conjugate transpose of G, and Σ is the DMFT self-energy, which requires a solution of the quantum impurity problem embedded in a self-consistent medium. We used a vertex-corrected one-crossing approximation (*7*), and the results were further cross-checked against a continuous-time quantum Monte Carlo method (*9*, *10*). The Slater integrals F^2, F^4, and F^6 were computed by the atomic physics program of (*11*), and F^0 was estimated by the constrained LDA to be 5 eV (*12*). The localized Ce 4*f* orbital was constructed from the nonorthogonal linear muffin-tin orbitals in a particular way to maximize its *f* character, as explained elsewhere (*13*).

The spectral function of *f* electron materials has been known to exhibit notable many-body effects. Setting the stage for their theoretical description, Fig. 1B displays the Ce 4*f* local spectral function [i.e., $A(\omega) = \Sigma_{\mathbf{k}} A(\mathbf{k},\omega)$], which is measured in angle-integrated photoemission experiments.

At room temperature, there is very little spectral weight at the Fermi level because the *f* electrons are tightly bound and localized on the Ce atom, giving rise to a broad spectrum concentrated mainly in the lower and upper Hubbard bands at –2.5 and +3 eV, respectively.

As the temperature is decreased, a narrow peak appears near the Fermi level (Fig. 1B). The states forming this peak have a small but finite dispersion, and therefore the area of the peak can be interpreted as the degree of *f* electron delocalization. This quantity, as well as the scattering rate of the Ce 4*f* states [Im $\Sigma(\omega = 0)$], exhibit a clear crossover at a temperature scale T^* of the order of 50 K (Fig. 1C).

Our results are consistent with the angle-integrated photoemission measurements (*14*) in which the onset of states with *f* character at the Fermi level was observed. But the experimental resolution has to be improved by one order of magnitude to resolve the narrow peak predicted by the theory.

We now turn to the total (traced over all orbitals) momentum-resolved spectral function $\mathrm{Tr}[A(\mathbf{k},\omega)]$ plotted along symmetry directions in the Brillouin zone. In a band-theory description, it would be sharply peaked on a series of bands $\varepsilon_n(\mathbf{k})$, and the weight of those peaks would be unity. It is worthwhile to compare the high-intensity features of the LDA+DMFT spectra (color coded) with the LDA bands [$\varepsilon_n(\mathbf{k})$, shown as blue lines] (Fig. 2A). In the region below –1 eV, there is a good correspondence between them. However, there is the systematic downshift (indicated by a green arrow) of the LDA+DMFT features relative to the LDA bands (which have mainly In 5*p* and Ir 5*d* character). Unexpectedly, a similar trend is seen in the ARPES of (*15*), which we redraw in Fig. 2B. The position of these bands is weakly temperature dependent and, if warmed to room temperature, an almost rigid upward shift of 5 meV was identified in our theoretical treatment. Experimentally, it was not possible to resolve the momentum in the *z* direction; therefore, the same experimental data (which can be thought as the average of the two paths from Γ to *X* and from *Z* to *R*), where Γ, *X*, *Z*, and *R* are symmetry points in the momentum space shown in Fig. 2B, are repeated in the two directions.

Near the Fermi level (between –0.5 and +1 eV), there are substantial discrepancies between the LDA bands (which, in this region, have substantial *f* character) and the LDA+DMFT features. The correlations treated by LDA+DMFT substantially modify the spectral function features with *f* content, transferring spectral weight into the upper Hubbard band located around +3 eV (white region in Fig. 2A).

Hubbard bands are excitations localized in real space, without a well-defined momentum, and therefore they show up as a blurred region of spectral weight in the momentum plot of Fig. 2A. There is also a lower Hubbard band around –2.5 eV, which is hardly detectable in this figure. The reason for this is that it carries a very small spectral weight, which is redistributed over a broad frequency region as shown in Fig. 1B.

It is also useful to compare the LDA+DMFT Hubbard bands with those obtained with the more familiar LDA+U method, which is equivalent to the Hartree-Fock approximation of the LDA+DMFT method. The LDA+U method inserts a sharp nondispersive band around

–2.5 eV and substantially twists the rest of the conduction bands. For the purpose of describing the set of bands below –1 eV, the LDA+DMFT method is therefore closer to the LDA type of calculation with *f* bands removed from the valence band.

To obtain further insights into the nature of the low-energy spectra, we show the momentum-resolved *f* electron spectral function of Fig. 1B in Fig. 2, C and D. The two plots correspond to the low-temperature (10 K) and high-temperature (300 K) spectra, respectively. At room temperature, a set of broad and dispersive bands is seen and should be interpreted as the *spd* bands leaving an imprint in the *f* electron spectral function resulting from hybridization. At low temperature, a narrow stripe of spectral weight appears at zero frequency, which cuts the conduction bands and splits them into two separate pieces, divided by a new hybridization gap.

The two straight nondispersive bands at –0.3 and +0.3 eV can also be identified in Fig. 2C and are due to the spin-orbit coupling (*16*). The same splitting of the coherence peak can be identified in the local spectra plotted in Fig. 1B and was recently observed in an ARPES study (*14*).

A detailed analysis of the zero-energy stripe of spectra in Fig. 2C reveals that the low-energy features correspond to the three very narrow bands (the dispersion is of the order of 3 meV) crossing the Fermi level. This is the origin of the large effective mass and large specific heat of the material at low temperatures. The low-energy band structure and its temperature dependence are theoretical predictions that can be verified experimentally in future ARPES studies.

Optical conductivity is a very sensitive probe of the electronic structure and has been applied to numerous heavy fermion materials (*17*). It is a technique that is largely complementary to the photoemission on two counts: It probes the bulk (and not the surface), and it is most sensitive to the itinerant *spd* electrons (rather than the *f* electrons).

A prototypical heavy fermion at high temperatures has an optical conductivity characterized by a very broad Drude peak. At low temperatures, optical data are usually modeled in terms of transitions between two renormalized bands, separated by a hybridization gap. These two bands give rise to a very narrow Drude peak of small weight and an optical absorption feature above the hybridization gap, termed the mid-infrared peak. This picture qualitatively describes the experimental data of $CeIrIn_5$ (*18*, *19*), which we reproduce in Fig. 3B. However, this simplified two-band model fails to account for some aspects of the data. For example, at 10 K, there is a clear structure in the mid-infrared peak. In addition to the broad shoulder around 0.07 eV, a second peak around 0.03 eV can be

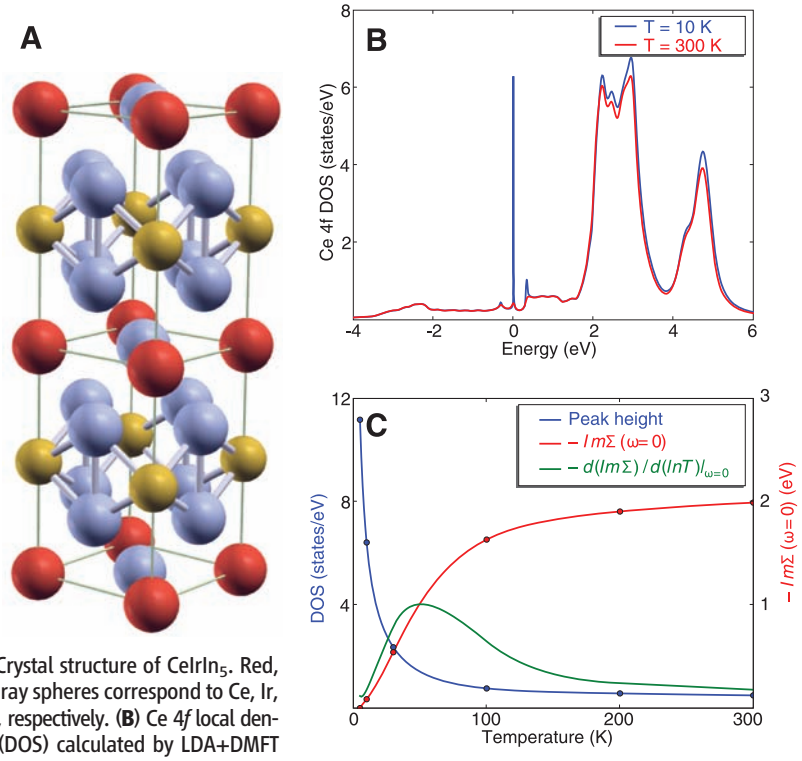


Fig. 1. (**A**) Crystal structure of $CeIrIn_5$. Red, yellow, and gray spheres correspond to Ce, Ir, and In atoms, respectively. (**B**) Ce 4*f* local density of state (DOS) calculated by LDA+DMFT at 10 and 300 K. (**C**) The quasiparticle peak height versus temperature (blue points), the imaginary part of Ce $4f_{5/2}$ self-energy $\Sigma_f(\omega = 0)$ (red points), and its temperature derivative (green points). The buildup of coherence is very slow and gradual. Around $T^* \sim 50$ K, the coherence first sets in and manifests itself in a fast increase of the quasiparticle peak (blue line) and crossover in the scattering rate [the derivative of scattering rate (green line) is peaked at T^*]. The quasiparticle peak weight saturates at much lower temperature (<5 K) and drops to zero at very high temperature, displaying a very long logarithmic tail.

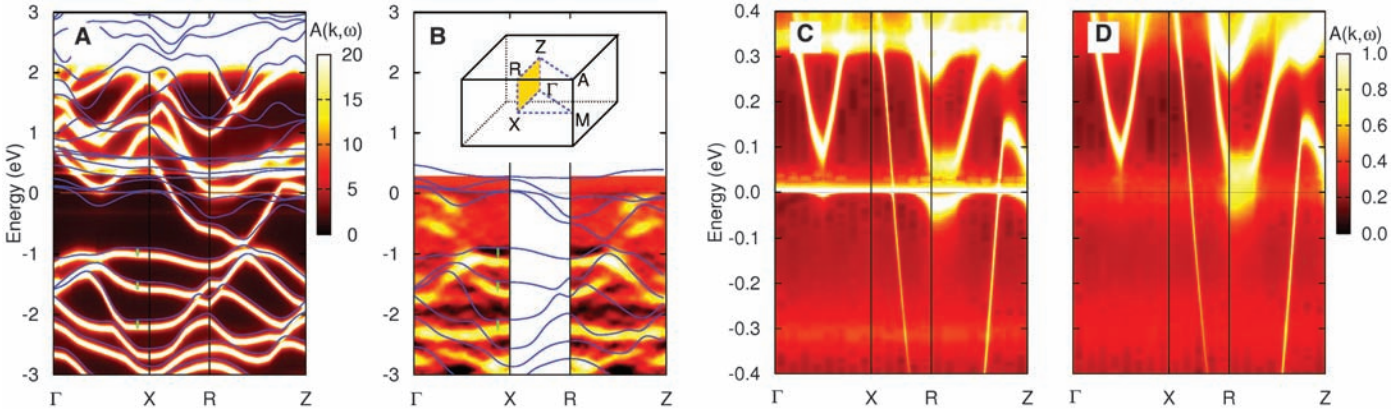


Fig. 2. (**A**) Momentum-resolved total spectral functions calculated by the LDA+DMFT method at 10 K are shown by color scheme. The LDA bands are indicated as blue lines. (**B**) Color plot shows experimental ARPES data reproduced from (*15*). (**C** and **D**) The momentum-resolved Ce 4*f* spectral function at 10 K (C) and 300 K (D).

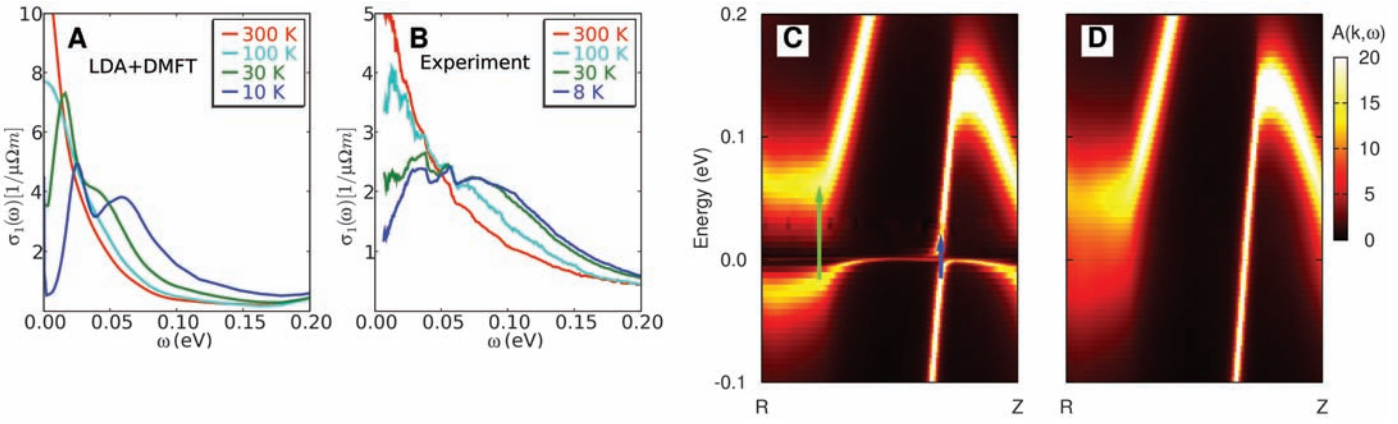


Fig. 3. The optical conductivity at several temperatures (**A**) obtained by LDA+DMFT and (**B**) measured experimentally and reproduced from (*18*). (**C** and **D**) The momentum-resolved non–Ce 4*f* spectral function ($A_{total} - A_{Ce\ 4f}$) at 10 K (C) and 300 K (D). Blue arrow, 30-meV hybridization gap; green arrow, 70-meV hybridization gap.

identified, which was previously interpreted as the absorption on the bosonic mode that might bind electrons in the unconventional superconducting state (*18*, *20*).

Also, the hybridization gap in simplified theories gives rise to a sharp drop of conductivity below the energy of the gap, whereas broader features are seen experimentally.

Optical conductivity within LDA+DMFT was recently implemented (*21*), and we show the results in Fig. 3A. They bear a strong resemblance to the experimental data in Fig. 3B. For example, a broad Drude peak at high temperature and a very clear splitting of the mid-infrared peak at low temperature are visible.

To understand the physical origin of these multiple peaks, we plot the momentum-resolved conduction electron (non–Ce 4*f*) spectral function along a representative high-symmetry line at 10 and 300 K, respectively (Fig. 3, C and D). There is a dramatic difference between the two temperatures. At high temperature, we see two bands in this momentum direction: One has a very short lifetime and the other has a long lifetime. The spread of the left band in Fig. 3D is due to electron-electron scattering, which broadens the band for about 100 meV. The character of both bands is primarily of In 5*p*, with an important difference. The left band comes mostly from the In atoms in the $IrIn_2$ layer, whereas the right band is mostly due to In in the $CeIn_3$ layer. The latter In atoms will be called in plane (each Ce atom has four neighbors of this type), and the former will be called out of plane (there are eight nearest neighbors to the Ce atom). As the temperature is lowered, the two In bands hybridize in a very different way with the Ce 4*f* moment. It is quite unexpected that the in-plane In atoms hybridize less with the Ce moment, leading to a small hybridization gap of the magnitude 30 meV (blue arrow in Fig. 3C). The out-of-plane In atoms are more coupled to the Ce moment, which leads to a larger hybridization gap of the order of 70 meV (green arrow in Fig. 3C). The existence of multiple hybridization gaps results in the splitting of the mid-infrared peak in optical conductivity (Fig. 3A).

The fact that the Ce moment is more coupled to out-of-plane In than to in-plane In provides a natural explanation for why these materials are sensitive to the substitution of transition metal ion Ir with Co or Rh. Namely, the out-of-plane In atoms are strongly coupled not only to Ce but also to the transition metal ion in their immediate neighborhood, whereas the in-plane In atoms are insensitive to the substitution.

Some of the results of the microscopic theory, such as the momentum-dependent hybridization (*19*) and the slow buildup of coherence (*22*), were foreshadowed by earlier phenomenological approaches. The first-principles DMFT treatment places these ideas within a microscopic framework.

In investigating the formation of the heavy fermion state with temperature in $CeIrIn_5$, we have shown that incorporating local correlations, only on the *f* site, allows for a coherent description of the evolution of the one-electron spectra and the optical conductivity with temperature. The approach provides a natural explanation for many unexpected features observed in this material and makes several quantitative predictions for the evolution of the spectra as a function of temperature, which can be tested by ARPES measurements currently under way. Although the single-site DMFT description is sufficient in a broad region of temperatures and parameters, cluster extensions of DMFT will be necessary to address the quantum criticality that takes place as Ir is replaced by Rh and Co and to address the possible instabilities toward unconventional superconductivity. Although model cluster DMFT studies seem very promising, the implementation of these methods, in conjunction with realistic electronic structure, remains a challenge for the future. Furthermore, in the treatment of other compounds of the same class (Ir substituted by Co or Rh), the correlations on the 3*d* or 4*d* transition metal will require GW approximation to treat the electronic structure.

References and Notes

1. G. R. Stewart, *Rev. Mod. Phys.* **56**, 755 (1984).
2. J. W. Allen, *J. Phys. Soc. Jpn.* **74**, 34 (2005).
3. C. Petrovic *et al.*, *Europhys. Lett.* **53**, 354 (2001).
4. Y. N. Grin, Y. P. Yarmolyuk, E. I. Giadyshevskii, *Sov. Phys. Crystallogr.* **24**, 137 (1979).
5. E. G. Moshopoulou, Z. Fisk, J. L. Sarrao, J. D. Thompson, *J. Solid State Chem.* **158**, 25 (2001).
6. G. Kotliar, D. Vollhardt, *Phys. Today* **57**, 53 (2004).
7. G. Kotliar *et al.*, *Rev. Mod. Phys.* **78**, 865 (2006).
8. S. Y. Savrasov, *Phys. Rev. B* **54**, 16470 (1996).
9. P. Werner, A. Comanac, L. de' Medici, M. Troyer, A. J. Millis, *Phys. Rev. Lett.* **97**, 076405 (2006).
10. K. Haule, *Phys. Rev. B* **75**, 155113 (2007).
11. R. D. Cowan, *The Theory of Atomic Structure and Spectra* (Univ. of California Press, Berkeley, CA, 1981).
12. A. K. McMahan, C. Huscroft, R. T. Scalettar, E. L. Pollock, *J. Comput. Aided Mater. Des.* **5**, 131 (1998).
13. A. Toropova, C. A. Marianetti, K. Haule, G. Kotliar, *Phys. Rev. B* **76**, 155126 (2007).
14. S. I. Fujimori *et al.*, *Phys. Rev. B* **73**, 224517 (2006).
15. S. I. Fujimori *et al.*, *Phys. Rev. B* **67**, 144507 (2003).
16. A. Sekiyama *et al.*, *Nature* **403**, 396 (2000).
17. L. Degiorgi, *Rev. Mod. Phys.* **71**, 687 (1999).
18. F. P. Mena, D. van der Marel, J. L. Sarrao, *Phys. Rev. B* **72**, 045119 (2005).
19. K. S. Burch *et al.*, *Phys. Rev. B* **75**, 054523 (2007).
20. E. J. Singley, D. N. Basov, E. D. Bauer, M. B. Maple, *Phys. Rev. B* **65**, 161101 (2002).
21. K. Haule, V. Oudovenko, S. Y. Savrasov, G. Kotliar, *Phys. Rev. Lett.* **94**, 036401 (2005).
22. S. Nakatsuji, D. Pines, Z. Fisk, *Phys. Rev. Lett.* **92**, 016401 (2004).

23. We thank S. I. Fujimori and A. Fujimori for their published data of ARPES on $CeIrIn_5$; F. P. Mena, D. van der Marel, and J. L. Sarrao for their published data of optics on $CeIrIn_5$; and J. W. Allen for unpublished data on other heavy fermion compounds. Work supported by the NSF Division of Material Research (grant no. 0528969).

10 August 2007; accepted 16 October 2007
Published online 1 November 2007;
10.1126/science.1149064
Include this information when citing this paper.

Designing Superoleophobic Surfaces

Anish Tuteja,[1] Wonjae Choi,[2] Minglin Ma,[1] Joseph M. Mabry,[3] Sarah A. Mazzella,[3] Gregory C. Rutledge,[1] Gareth H. McKinley,[2]* Robert E. Cohen[1]*

Understanding the complementary roles of surface energy and roughness on natural nonwetting surfaces has led to the development of a number of biomimetic superhydrophobic surfaces, which exhibit apparent contact angles with water greater than 150 degrees and low contact angle hysteresis. However, superoleophobic surfaces—those that display contact angles greater than 150 degrees with organic liquids having appreciably lower surface tensions than that of water—are extremely rare. Calculations suggest that creating such a surface would require a surface energy lower than that of any known material. We show how a third factor, re-entrant surface curvature, in conjunction with chemical composition and roughened texture, can be used to design surfaces that display extreme resistance to wetting from a number of liquids with low surface tension, including alkanes such as decane and octane.

Many surfaces in nature, including various plant leaves (*1*, *2*), legs of the water strider (*3*), troughs on the elytra of desert beetles (*4*), and geckos' feet (*5*, *6*), are superhydrophobic, displaying apparent contact angles with water (surface tension γ_{lv} = 72.1 mN/m) greater than 150° and low contact angle hysteresis. Understanding the complementary roles of the two key surface parameters, surface energy and roughness (*7*–*11*), for these materials has led to the development of a number of artificial superhydrophobic surfaces (*6*, *12*, *13*). However, superoleophobic surfaces—structured surfaces that resist wetting of liquids with much lower surface tension, such as decane (γ_{lv} = 23.8 mN/m) or octane (γ_{lv} = 21.6 mN/m)—are extremely rare (*14*). We have developed several different textured surfaces displaying contact angles greater than 160°, even with octane. This observed superoleophobicity can be explained by considering local surface curvature as the third parameter that affects both the apparent contact angle and hysteresis on any surface. This understanding also allows us to rationalize numerous earlier observations of unexpectedly high liquid repellency on rough surfaces (*14*–*18*).

The best-known example of a natural superhydrophobic surface is the surface of the lotus leaf, *Nelumbo nucifera* (*1*). Numerous studies have suggested that the superhydrophobic character of the lotus leaf surface is attributable to a combination of surface chemistry and roughness on multiple scales. However, a liquid with a markedly lower surface tension, such as hexadecane (γ_{lv} = 27.5 mN/m), spreads rapidly across the lotus leaf, leading to a contact angle of ~0° (fig. S8) (*19*). Such surface oleophilicity can have an impact on a wide range of phenomena, including biofouling by marine organisms, loss of self-cleaning ability of plant leaves in polluted waters, and swelling of elastomeric seals and O-rings.

Two distinct models, developed independently by Wenzel (*7*) and Cassie and Baxter [(*8*), henceforth the Cassie model], are commonly used to explain the effect of roughness on the apparent contact angle of liquid drops (*20*, *21*). The Wenzel model recognizes that surface roughness increases the available surface area of the solid, which modifies the surface contact angle according to the expression

$$\cos\theta^* = r\cos\theta \qquad (1)$$

where θ^* is the apparent contact angle on the textured surface, r is the surface roughness, and θ is the equilibrium contact angle on a smooth surface of the same material, given by Young's equation (*22*) as $\cos\theta = (\gamma_{sv} - \gamma_{sl})/\gamma_{lv}$, where γ refers to the interfacial tension and the subscripts s, l, and v refer to the solid, liquid, and vapor phases, respectively.

The Cassie model, on the other hand, postulates that the superhydrophobic nature of a rough surface is caused by microscopic pockets of air remaining trapped below the liquid droplet, leading to a composite interface. If ϕ_s is the fraction of the solid in contact with the liquid, the Cassie equation yields

$$\cos\theta^* = -1 + \phi_s(1 + \cos\theta) \qquad (2)$$

In contrast to the Wenzel relation, the Cassie relation allows for the possibility of $\theta^* > 90°$, even with $\theta < 90°$. Thermodynamic considerations can be used to determine whether a particular textured surface will exist in the Wenzel or the Cassie state [see recent work by Marmur (*10*) and Nosonovsky (*11*)]. The situation is somewhat complicated by the presence of multiple local free energy minima leading to so-called "metastable" configurations (*2*, *10*, *23*, *24*). Indeed, Krupenkin *et al*. (*25*) have highlighted the possibility of transitioning reversibly between the Wenzel and Cassie states (*19*). However, careful experimentation with model microstructured surfaces and corresponding free energy calculations show that a series of rough substrates with progressively increasing equilibrium contact angles exhibits a transition from the Wenzel to the Cassie state (*24*, *26*, *27*). The threshold value of the equilibrium contact angle (θ_c) for this transition is obtained by equating Eqs. 1 and 2:

$$\cos\theta_c = (\phi_s - 1)/(r - \phi_s) \qquad (3)$$

Because $r > 1 > \phi_s$, the critical value of the equilibrium contact angle θ_c for this transition is necessarily greater than 90°. Thus, it may be anticipated that for $\theta < 90°$, a surface cannot exist in the Cassie state or that the creation of highly nonwetting surfaces ($\theta^* >> 90°$) requires $\theta > \theta_c > 90°$. These arguments highlight the difficulty of developing surfaces for which $\theta^* > 150°$ when in contact with alkanes such as decane or octane, as there are no reports of a natural or artificial surface with a low enough surface energy (*19*) to enable $\theta > 90°$ with these liquids (*14*–*17*, *28*).

However, studies on leaves of plants such as *Cotinus coggygria* and *Ginkgo biloba* suggest the possibility of designing textured surfaces with unexpectedly high nonwetting properties. Herminghaus (*2*) first pointed out that these leaves display superhydrophobic properties, even with $\theta < 90°$. Indeed, recent experiments show that even the wax on the lotus leaf surface is weakly hydrophilic with $\theta \approx 74°$ (*29*). The superhydrophobic state of these textured surfaces is not the true equilibrium state, and submerging the leaf in water to a certain depth can cause a transition from this metastable Cassie state to the Wenzel state. Correspondingly, it may be possible to design metastable superoleophobic surfaces even though we are limited to materials with $\theta < 90°$.

We have synthesized a class of hydrophobic polyhedral oligomeric silsesquioxane (POSS) molecules in which the rigid silsesquioxane cage is surrounded by perfluoro-alkyl groups (inset, Fig. 1A). A number of molecules with different organic groups including 1*H*,1*H*,2*H*,2*H*-heptadecafluorodecyl (referred to as fluorodecyl POSS) and 1*H*,1*H*,2*H*,2*H*-tridecafluorooctyl (fluorooctyl POSS) were synthesized. We refer to this class of materials generically as fluoroPOSS (*19*). The high surface concentration and surface mobility of $-CF_2$ and $-CF_3$ groups, together with the relatively high ratio of $-CF_3$ groups with respect to $-CF_2$ groups, results in a very hydrophobic material with low surface energy. A film of fluorodecyl POSS, spin-coated on a Si wafer, has an advancing (θ_{adv}) and receding (θ_{rec}) contact angle of 124.5° ± 1.2°, with a root mean square (RMS) roughness r_q = 3.5 nm.

By varying the mass fraction of specific fluoroPOSS molecules blended with a polymer, we can systematically change γ_{sv} for the polymer-fluoroPOSS blend. We studied blends of a moderately hydrophilic polymer, poly(methyl methacrylate) (PMMA), and fluorodecyl POSS (which had the lowest surface energy of the POSS molecules synthesized). Varying the mass fraction of fluo-

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Air Force Research Laboratory, Edwards Air Force Base, CA 93524, USA.

*To whom correspondence should be addressed. E-mail: gareth@mit.edu (G.H.M.); recohen@mit.edu (R.E.C.)

rodecyl POSS systematically changed both θ_{adv} and θ_{rec} for water on spin-coated films of PMMA and fluorodecyl POSS (Fig. 1A). The atomic force microscope (AFM) phase images of the spin-coated surfaces, as well as the shapes of water droplets (volume ≈ 2 μl) on those surfaces, are shown in Fig. 1B. Comparing the phase images of the pure PMMA and 1.9 weight % fluorodecyl POSS indicates substantial surface migration of POSS during solvent evaporation, as would be expected because of its low surface energy (*30*, *31*).

The corresponding rough fluorodecyl POSS–PMMA surfaces are created by electrospinning (*32*–*34*). In Fig. 1C we show the advancing (θ^*_{adv}) and receding θ^*_{rec} contact angles with water for the various electrospun surfaces. The inset shows the "beads on a string" (*35*) morphology of a representative fiber mat, as well as the multiple scales of roughness and high porosity generated by the electrospinning process. There is no observable change in the micrometer-scale structure with increasing mass fraction of POSS as viewed with a scanning electron microscope (SEM). X-ray photoelectron spectroscopy (XPS) analysis indicates substantial surface migration of the fluoroPOSS molecules during electrospinning (fig. S9). The surfaces become superhydrophobic (movie S1) for all POSS concentrations above ~10 weight % ($\theta^*_{adv} = \theta^*_{rec} = 161° \pm 2°$).

An important observation from Fig. 1C is that $\theta^*_{adv} > 90°$ for pure PMMA and 1.9 weight % POSS–PMMA electrospun surfaces, even though in each case $\theta_{adv} < 90°$ (Fig. 1A). This is surprising, because for $\theta < 90°$ the rough surfaces are expected to be in the Wenzel state, and from Eq. 1 we expect $\theta^*_{adv} < \theta_{adv}$. This unusual effect is further explored in the form of the general wetting diagram (*13*) in Fig. 1D, which shows a plot of cos θ^*_{adv} and cos θ^*_{rec} on the rough electrospun surfaces as a function of cos θ_{adv} and cos θ_{rec} for the corresponding smooth (spin-coated) surfaces. Even at low POSS concentrations (<2 weight %), the surfaces display high apparent advancing contact angles indicative of being in the Cassie state. However, these textured surfaces also exhibit high contact angle hysteresis. Separate experiments show that this Cassie state is metastable, as water droplets released from a height can penetrate and wet the fiber mat.

The electrospun fibers similarly display extremely oleophobic properties ($\theta^* >> 90°$), even though all of the corresponding spin-coated surfaces are oleophilic ($\theta < 90°$). Plots of θ^*_{adv} and θ^*_{rec} for hexadecane and decane (Fig. 2, A and B) show that in many cases both θ^*_{adv} and θ^*_{rec} for the electrospun surfaces are much greater than 90°, and a transition from the Wenzel to the metastable Cassie state can also be observed for each alkane. This metastable state is related to a local minimum in free energy (*19*) that systematically shifts to higher POSS concentrations (lower surface energy) with decreasing liquid surface tension. The electrospinning process also enables fibers to be deposited on fragile or natural surfaces (such as a lotus leaf) to confer oleophobicity in addition to superhydrophobicity (fig. S8).

The advancing contact angles for various liquid alkanes on the electrospun and spin-coated surfaces can be combined to form a master curve of θ^*_{adv} versus θ_{adv} (Fig. 2C). A sharp transition from the Wenzel state to the nonwetting Cassie state is observed at an equilibrium contact angle of $\theta_c \approx 69°$ (see also fig. S10). This value of θ_c depends on the free

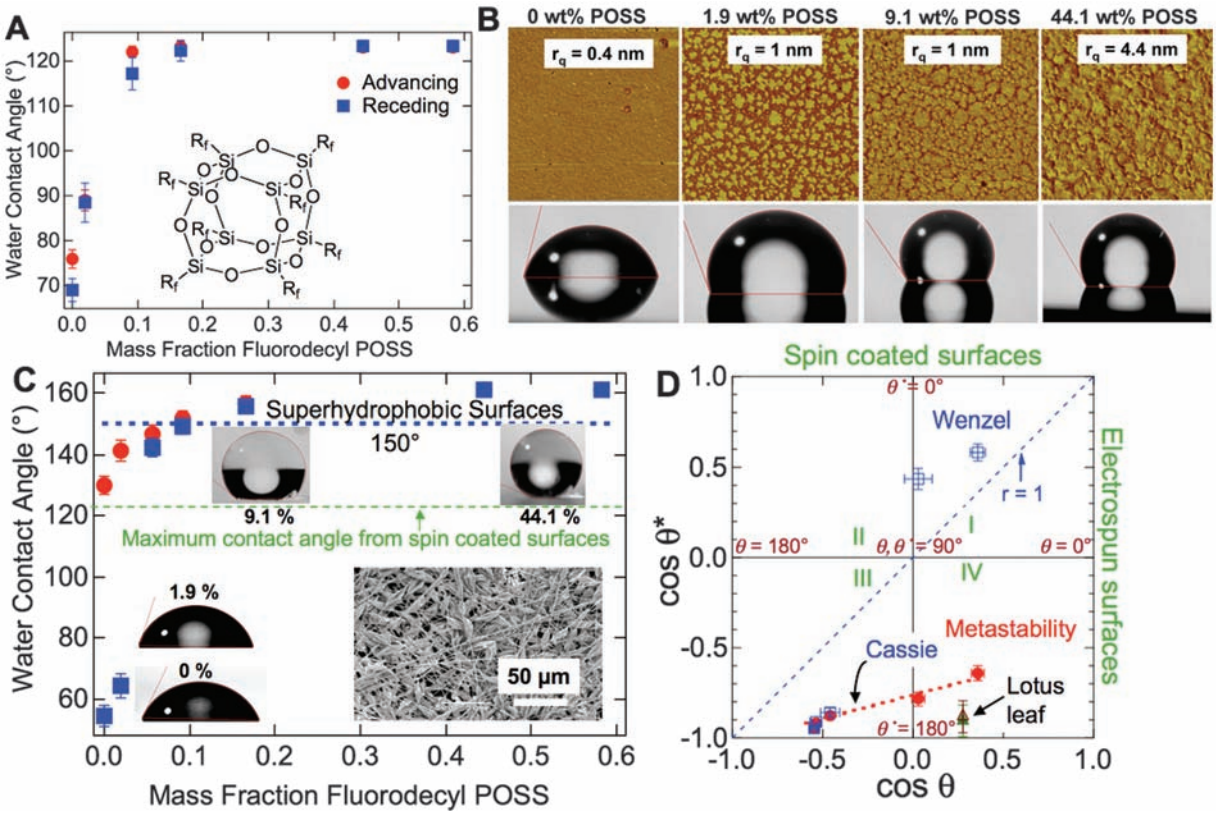


Fig. 1. Tunable hydrophobicity of fluoroPOSS-PMMA blends. (**A**) θ_{adv} and θ_{rec} for water as a function of the mass fraction of fluorodecyl POSS. The inset shows the general molecular structure of fluoroPOSS molecules. The alkyl chains (R_f) have the general molecular formula $-CH_2CH_2(CF_2)_nCF_3$, where n = 0, 3, 5, or 7. (**B**) The phase angle scale on the AFM images is 0° to 10° for the 0, 9.1, and 44.1 weight % POSS images and 0° to 90° for the 1.9 weight % POSS image. The RMS roughness (r_q) for each film is also given. By comparison, r_q for a Si wafer is ~0.2 nm. (**C**) θ^*_{adv} (red dots) and θ^*_{rec} (blue squares) for water on the electrospun surfaces. The inset shows a SEM micrograph for an electrospun surface containing 9.1 weight % POSS. The maximum contact angle for water on the spin-coated surfaces is also shown. (**D**) Plot of cos θ^*_{adv}(solid red dots) and cos θ^*_{rec} (open blue squares) for water as a function of θ_{adv} and θ_{rec}. The surfaces in the lower right quadrant (IV) of this diagram correspond to hydrophilic substrates that are rendered hydrophobic purely by topography. The advancing and receding contact angles for the lotus leaf (solid and open triangles, respectively) are also provided for comparison.

energy landscape separating the Cassie and Wenzel states (*19*) and will be dependent on the topography of the surfaces under consideration; however, it is clear from Fig. 2C and fig. S10 that θ_c can be considerably less than 90°.

The robustness of the metastable Cassie state for the electrospun fiber surfaces is determined by measuring the height of liquid (or static pressure) required for forcing the liquid through the fibers (denoted h^*). The surfaces do not transition to the Wenzel state even when submerged under 150 mm of hexadecane. To scale such observations, we use the characteristic capillary rise height

$$h_{cap} = 2\gamma_{lv} \cos\theta/\rho g D \qquad (4)$$

for the porous substrate, where the relevant capillary pore size $2D$ is the average edge-to-edge spacing of the nanofibers (Fig. 3A) (*8*), ρ is the fluid density, and g is the acceleration due to gravity. In Fig. 2D we plot the normalized breakthrough pressure

$$h^*/h_{cap} = \rho g D h^*/(2\gamma_{lv} \cos\theta) \qquad (5)$$

required to wet a fiber mat containing 44 weight % POSS. Calculations shown in (*19*) yield the pressure required to increase the curvature of a liquid droplet until it impinges on the next level of fibers. These pressures (normalized with the capillary pressure) are also shown in Fig. 2D.

An extremely useful application for these porous electrospun materials can be deduced by noticing that many of the electrospun surfaces with low POSS concentrations are both superhydrophobic and superoleophilic ($\theta^*_{alkane} \approx 0°$). Thus, these surfaces are ideal for separating mixtures or dispersions of alkanes (oils) and water (*36*), as shown in Fig. 2E.

It is the local surface curvature that plays a key role in driving the oleophobicity of electro-

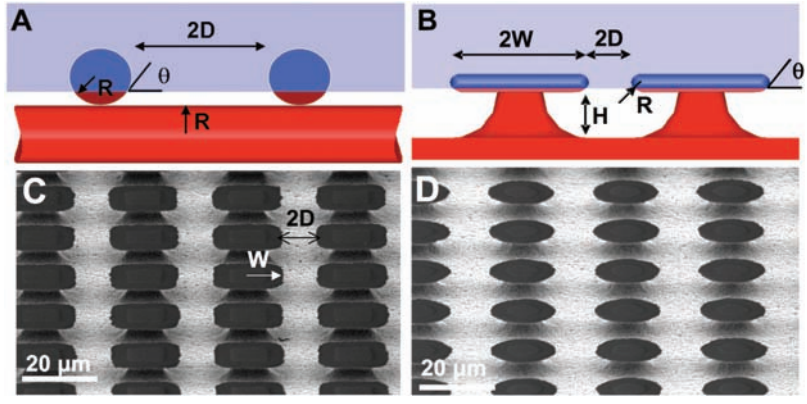


Fig. 3. Critical role of re-entrant curvature. (**A** and **B**) Cartoons highlighting the formation of a composite interface on surfaces with re-entrant topography (for both fibers and micro-hoodoos). The geometric parameters R, D, H, and W characterizing these surfaces are also shown. The blue surface is wetted while the red surface remains nonwetted when in contact with a liquid whose equilibrium contact angle is θ (< 90°). (**C** and **D**) SEM micrographs for two micro-hoodoo surfaces having square and circular flat caps, respectively. The samples are viewed from an oblique angle of 30°.

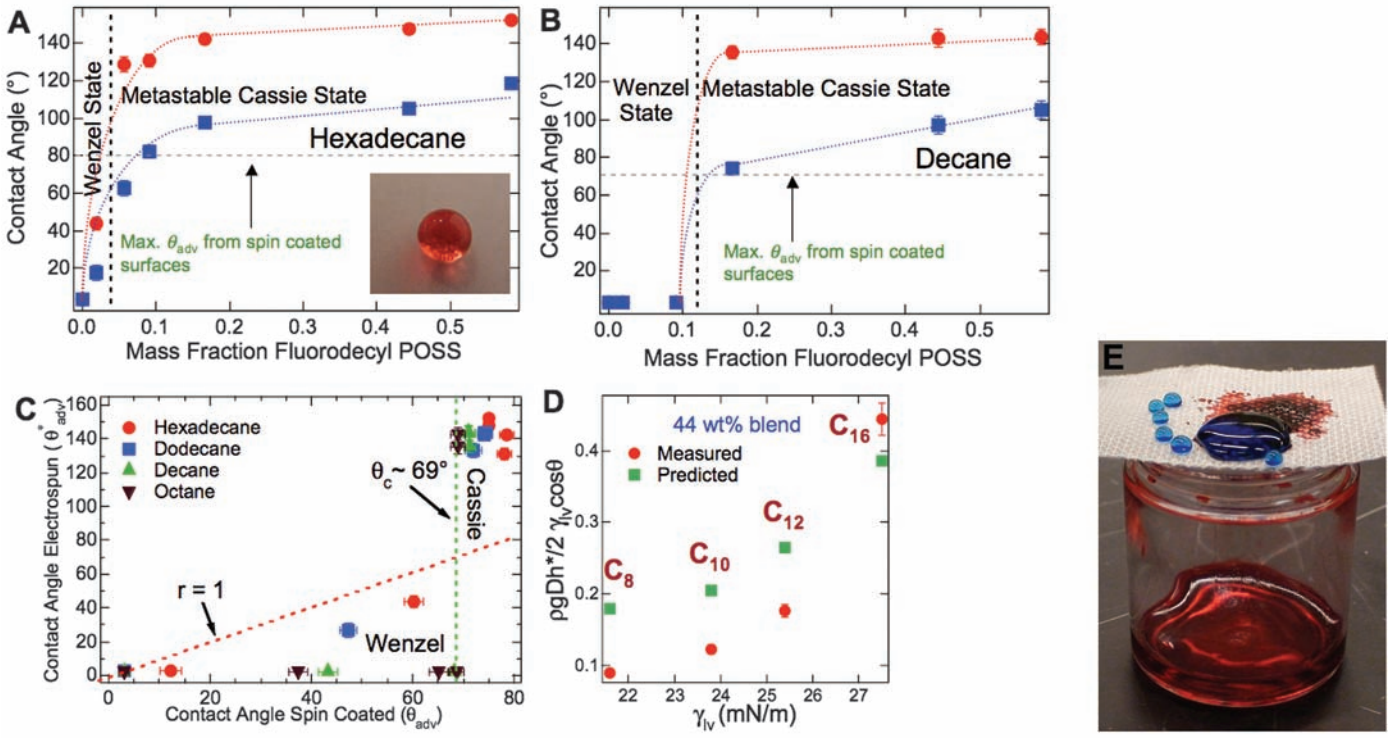


Fig. 2. (**A** and **B**) θ^*_{adv} (red dots) and θ^*_{rec} (blue squares) for hexadecane and decane, respectively, on the electrospun surfaces. The maximum contact angles on corresponding spin-coated surfaces (θ_{adv}) are also shown. The inset of (A) shows a drop of hexadecane (dyed with Oil Red O) on a 44 weight % fluorodecyl POSS electrospun surface. (**C**) θ^*_{adv} for hexadecane, dodecane, decane, and octane as a function of θ_{adv}. (**D**) Normalized breakthrough pressure required to transition irreversibly from the metastable Cassie to the Wenzel state on the surface of fibers containing 44 weight % fluorodecyl POSS. Our predictions for the breakthrough pressure (*19*) using R = 500 nm and D = 4 μm are also shown. (**E**) A steel grid (square pores with 1-mm spacing) coated with electrospun fibers containing 9.1 weight % fluorodecyl POSS used for oil-water separation. Octane droplets (colored with Oil Red O) easily pass through the membrane, whereas water droplets (dyed with methylene blue) bead up on the surface (see movie S2).

spun fibers. To support this thesis, we fabricated model re-entrant structures (i.e., surfaces having concave topographic features) of the form shown in Fig. 3, B to D. These structures were fabricated on flat Si wafers by means of SiO_2 deposition followed by a two-step etching process comprising reactive ion etching of SiO_2 and subsequent isotropic etching of Si with the use of vapor-phase XeF_2 (*19*). This results in undercut silicon pillars (covered with a 300-nm layer of SiO_2) and troughs. Because of their similarity to geomorphological features, we refer to these structures as micro-hoodoos (*37*). A key design feature of both of the electrospun fiber and hoodoo surfaces is that they possess re-entrant curvature, in addition to exhibiting both the desired characteristics of "roughness" (i.e., $r > 1$) and low wetted surface fraction (i.e., $\phi_s < 1$) embodied in Eqs. 1 and 2.

The flat liquid-air interface shown schematically in Fig. 3, A and B (*38*) is in contact with the electrospun fibers and the micro-hoodoos. As a direct result of re-entrant curvature, for any value of $0° < \theta < 180°$, each surface in Fig. 3, A and B, provides a point somewhere along the fiber or cap length at which Young's equation is satisfied locally at the air-liquid-solid interface (*10*, *11*, *34*). This is in contrast to other common patterned supports [e.g., arrays of vertical posts (*25*)], which can form a composite interface and satisfy the Young equation only if $\theta > 90°$. It is therefore possible to form a composite interface on re-entrant curved surfaces, with the drop sitting partially on air, and thus have $\theta^* >> 90°$ even with $\theta < 90°$. However, this Cassie state is only locally stable, because the total energy of the system decreases appreciably (*19*) if the liquid advances and completely covers the fibers/hoodoos, leading to a fully wetted interface (*11*, *39*).

Analysis of these re-entrant geometries yields two important design parameters (*19*). For the electrospun fibers, the first is the spacing ratio $D^* = (R + D)/R$ (where R is the radius of the fibers), which directly affects ϕ_s (*8*) and thus the apparent contact angle. The second is the robustness parameter $H^* = 2(1 - \cos\theta)Rl_{cap}/D^2$ [where $l_{cap} = (\gamma_{lv}/\rho g)^{1/2}$], which measures the robustness of the metastable Cassie state with respect to the fluid properties, equilibrium contact angle, and surface geometry. Varying the R and D values of these structures has competing effects on the apparent contact angles and the stability of the composite interface. Thus, the metastable Cassie state may not be accessible in practice on every surface that possesses re-entrant curvature, depending on the applied pressure required to transition irreversibly from the metastable Cassie to the Wenzel state. However, it is clear that to maximize both θ^* and the stability of the Cassie state, we seek sparsely spaced, highly re-entrant surfaces with both $H^* >> 1$ and $D^* >> 1$. This is why our electrospun mats of nanofibers ($D^* = 9$; $H^* = 46$ for octane on fibers containing 44.1 weight % POSS) provide a more robust Cassie state than do prototypical wire gratings (*8*) ($D^* = 1.9$ to 5.8; $H^* = 0.76$ to 7.6 for octane), even at higher porosities.

Evaluating the magnitudes of these dimensionless design parameters also helps explain the oleophobicity observed for aggregates of anodized alumina (*15*, *16*), plasma-treated cotton fibers (*14*), and spherical arrays of poly(tetrafluoroethylene) (PTFE) particles (*17*), as well as superhydrophobicity obtained on a hydrophilic substrate with a "popcorn-like" morphology (*18*). Re-entrant structures on the lotus leaf surface, in the form of "overhangs" (*2*, *40*) on the hemispherical nubs (inset, fig. S8A), are also the reason for its unexpectedly high hydrophobicity.

For the micro-hoodoo geometry, the design parameters take the form $D^* = [(W + D)/W]^2$ and $H^* = 2[(1 - \cos\theta)R + H]l_{cap}/D^2$. Because the hoodoo spacing (W) and height (H) can be varied independently for these surfaces (Fig. 3B), we can easily decouple variations in D^* and H^* and thus engineer surfaces with both higher apparent contact angles and much greater robustness than even the electrospun fibers can provide.

During the patterning and etching process for fabricating the micro-hoodoos, we also left untextured areas, thus enabling the measurement of both θ and θ^* on the same wafer. For water on the textured surfaces (Fig. 4A), re-entrant curvature in the absence of any chemical surface treatment leads to $\theta^*_{adv} \approx 143°$ ($\theta^*_{rec} \approx 104°$; $D^* = 2.3$). By comparison, for water on the smooth portion of the same wafer, $\theta_{adv} \approx 10°$. The robustness of the metastable Cassie state on the SiO_2 micro-hoodoo surface (with $H^* = 1560$ for water) is illustrated in Fig. 4B. The re-entrant surface resists both the advancing (*41*) and receding of the water droplet. The images of the receding droplet also highlight the high hysteresis ($\Delta\theta^* \approx 39°$) on the textured surface; however, it is clear that the water droplet has not reached the SiO_2 surface at the base of the hoodoos, in which case it would have been impossible to withdraw the water droplet completely.

The SiO_2 hoodoos were then treated with vapor-phase 1*H*,1*H*,2*H*,2*H*-perfluorodecyltrichlorosilane to lower γ_{sv} chemically and hence combine low surface energy with re-entrant surface curvature. The advancing and receding contact angles for octane on the silanized hoodoo surfaces ($D^* = 2.3$ to 9; $H^* = 64$ to 1040 for octane) are shown in Fig. 4C as a function of ϕ_s or $1/D^*$. The inset of Fig. 4C shows a drop of octane on a silanized micro-hoodoo surface ($\phi_s = 0.11$, $D^* = 9$, $H^* = 64$; $\theta^*_{adv} \approx 163°$, $\theta^*_{rec} \approx 145°$). A 10-μl drop of octane can easily be rolled off this surface by tilting to 15°. Corresponding measurements for octane on a smooth SiO_2 surface ($r_q \approx 0.5$ nm) covered with the same silane coating yielded only $\theta_{adv} \approx 55°$, $\theta_{rec} \approx 50°$. The contact angles monotonically increased with decreasing wetted area fraction ϕ_s. This is in broad agreement with Eq. 2 (*42*) and results from the variable resistance offered to the receding meniscus, which is expected to be proportional to the total number of re-entrant structural elements at the air-liquid-solid three-phase contact line. Silanized micro-hoodoos with $\phi_s = 0.03$ ($D^* = 36$, $H^* = 10$) are fully wetted by octane.

We have demonstrated two different approaches for fabricating surfaces possessing re-entrant curvature. In each case, the re-entrant textures allow for the possibility of constructing extremely nonwetting surfaces that can support

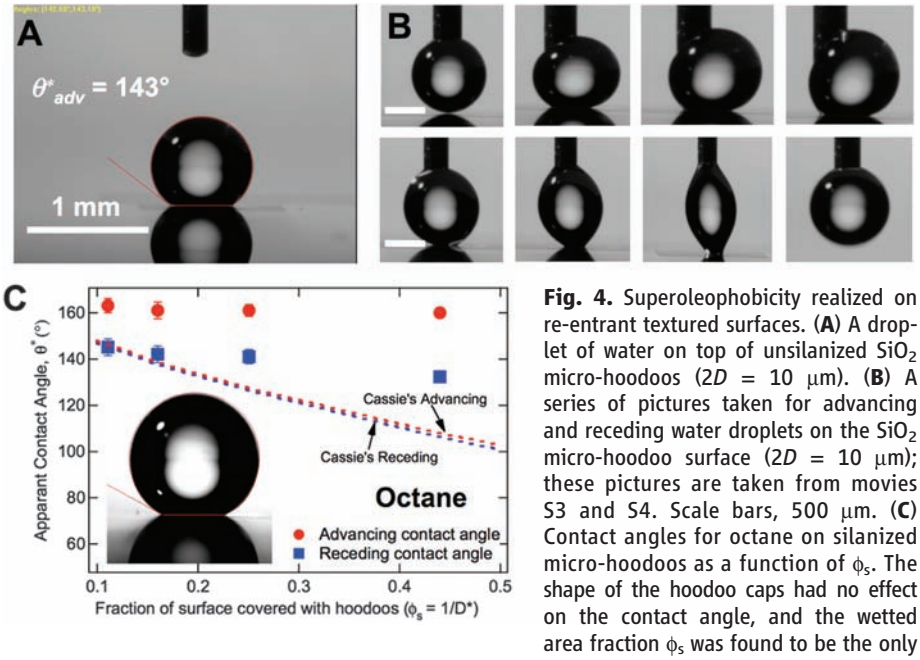


Fig. 4. Superoleophobicity realized on re-entrant textured surfaces. (**A**) A droplet of water on top of unsilanized SiO_2 micro-hoodoos ($2D$ = 10 μm). (**B**) A series of pictures taken for advancing and receding water droplets on the SiO_2 micro-hoodoo surface ($2D$ = 10 μm); these pictures are taken from movies S3 and S4. Scale bars, 500 μm. (**C**) Contact angles for octane on silanized micro-hoodoos as a function of ϕ_s. The shape of the hoodoo caps had no effect on the contact angle, and the wetted area fraction ϕ_s was found to be the only important parameter. Predictions from the Cassie equation (*8*) are also included for comparison (dashed lines).

the Cassie state with water and various organic liquids. The presence of re-entrant curvature, though, is not a sufficient condition for developing highly nonwetting surfaces; the Cassie state may be inaccessible in practice if the applied pressure (or energy barrier) required to transition from the Cassie to the Wenzel state is small. However, by independently controlling both the chemical and topographic nature of surfaces (as embodied in two dimensionless design parameters, D^* and H^*), we have shown that it is possible to design extremely robust nonwetting surfaces.

References and Notes

1. W. Barthlott, C. Neinhuis, *Planta* **202**, 1 (1997).
2. S. Herminghaus, *Europhys. Lett.* **52**, 165 (2000).
3. X. Gao, L. Jiang, *Nature* **432**, 36 (2004).
4. A. R. Parker, C. R. Lawrence, *Nature* **414**, 33 (2001).
5. K. Autumn *et al*., *Nature* **405**, 681 (2000).
6. J. Genzer, K. Efimenko, *Biofouling* **22**, 339 (2006).
7. R. N. Wenzel, *Ind. Eng. Chem.* **28**, 988 (1936).
8. A. B. D. Cassie, S. Baxter, *Trans. Faraday Soc.* **40**, 546 (1944).
9. M. Callies, D. Quere, *Soft Mat.* **1**, 55 (2005).
10. A. Marmur, *Langmuir* **19**, 8343 (2003).
11. M. Nosonovsky, *Langmuir* **23**, 3157 (2007).
12. A. Nakajima, K. Hashimoto, T. Watanabe, *Monatsh. Chem.* **132**, 31 (2001).
13. D. Quere, *Rep. Prog. Phys.* **68**, 2495 (2005).
14. S. R. Coulson, I. S. Woodward, J. P. S. Badyal, S. A. Brewer, C. Willis, *Chem. Mater.* **12**, 2031 (2000).
15. K. Tsujii, T. Yamamoto, T. Onda, S. Shibuichi, *Angew. Chem. Int. Ed. Engl.* **36**, 1011 (1997).
16. S. Shibuichi, T. Yamamoto, T. Onda, K. Tsujii, *J. Colloid Interface Sci.* **208**, 287 (1998).
17. W. Chen *et al*., *Langmuir* **15**, 3395 (1999).
18. M. Zhu, W. Zuo, H. Yu, W. Yang, Y. Chen, *J. Mater. Sci.* **41**, 3793 (2006).
19. See supporting material on *Science* Online.
20. L. Gao, T. J. McCarthy, *Langmuir* **23**, 3762 (2007).
21. Recent work by McCarthy and colleagues (*20*) points out that these models are applicable only at the solid-liquid-vapor three-phase contact line (drop perimeter) and that the interfacial area within the drop perimeter does not affect either the apparent contact angle or the hysteresis; thus, these models can be easily applied only to surfaces with a homogeneous texture, as considered here.
22. T. Young, *Philos. Trans. R. Soc. London* **95**, 65 (1805).
23. N. A. Patankar, *Langmuir* **19**, 1249 (2003).
24. B. He, N. A. Patankar, J. Lee, *Langmuir* **19**, 4999 (2003).
25. T. N. Krupenkin *et al*., *Langmuir* **23**, 9128 (2007).
26. L. Barbieri, E. Wagner, P. Hoffmann, *Langmuir* **23**, 1723 (2007).
27. A. Lafuma, D. Quere, *Nat. Mater.* **2**, 457 (2003).
28. W. A. Zisman, in *Contact Angle, Wettability and Adhesion*, F. M. Fowkes, Ed. (American Chemical Society, Washington, DC, 1964), pp. 1–51.
29. Y.-T. Cheng, D. E. Rodak, *Appl. Phys. Lett.* **86**, 144101 (2005).
30. D. E. Suk *et al*., *Macromolecules* **35**, 3017 (2002).
31. D. H. K. Pan, W. M. Prest Jr., *J. Appl. Phys.* **58**, 2861 (1985).
32. D. H. Reneker, A. L. Yarin, H. Fong, S. Koombhongse, *J. Appl. Phys.* **87**, 4531 (2000).
33. M. Ma, Y. Mao, M. Gupta, K. K. Gleason, G. C. Rutledge, *Macromolecules* **38**, 9742 (2005).
34. M. Ma *et al*., *Adv. Mater.* **19**, 255 (2007).
35. H. Fong, I. Chun, D. H. Reneker, *Polym.* **40**, 4585 (1999).
36. L. Feng *et al*., *Angew. Chem.* **116**, 2046 (2004).
37. These surfaces are referred to as micro-hoodoos because their geometry and process of creation are reminiscent of geological features called hoodoos, which are created by soil erosion. Hoodoos are composed of a soft sedimentary rock topped by a piece of harder, less easily eroded stone.
38. Because $2D << (\gamma_{lv}/\rho g)^{0.5}$ (capillary length; ρ is the density of liquid), the effect of gravity is negligible and we approximate the liquid-air interface to be a horizontal plane.
39. Nosonovsky (*11*) recently derived another important criterion for the creation of a local minimum in free energy, and thus for the creation of a stable heterogeneous interface: $dA_{sl}d\theta < 0$, where dA_{sl} is the change in solid-liquid contact area with the advancing or receding of the liquid, and $d\theta$ is the change in local contact angle. This criterion also emphasizes the importance of re-entrant surfaces.
40. L. Cao, H. H. Hu, D. Gao, *Langmuir* **23**, 4310 (2007).
41. As the surface is pushed toward the water droplet, the droplet moves, and hence the normal force is not transferred perfectly.
42. The difference in contact angle values from the Cassie prediction is related to contact line pinning.

43. Supported by Air Force Research Laboratory contract FA9300-06M-T015 and Air Force Office of Scientific Research contracts FA9550-07-1-0272 and LRIR-92PL0COR, with additional student support provided by the NSF Nanoscale Interdisciplinary Research Team on Nanoscale Wetting (DMR-0303916). We thank the Institute for Soldier Nanotechnologies at MIT for the use of facilities.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1618/DC1
Materials and Methods
SOM Text
Figs. S1 to S10
Tables S1 to S3
References
Movies S1 to S4

25 July 2007; accepted 22 October 2007
10.1126/science.1148326

The Equatorial Ridges of Pan and Atlas: Terminal Accretionary Ornaments?

Sébastien Charnoz,[1]* André Brahic,[1] Peter C. Thomas,[2] Carolyn C. Porco[3]

In the outer regions of Saturn's main rings, strong tidal forces balance gravitational accretion processes. Thus, unusual phenomena may be expected there. The Cassini spacecraft has recently revealed the strange "flying saucer" shape of two small satellites, Pan and Atlas, located in this region, showing prominent equatorial ridges. The accretion of ring particles onto the equatorial surfaces of already-formed bodies embedded in the rings may explain the formation of the ridges. This ridge formation process is in good agreement with detailed Cassini images showing differences between rough polar and smooth equatorial terrains. We propose that Pan and Atlas ridges are kilometers-thick "ring-particle piles" formed after the satellites themselves and after the flattening of the rings but before the complete depletion of ring material from their surroundings.

In images sent by the Voyager spacecraft in the early 1980s, two small satellites were discovered orbiting inside Saturn's rings (*1*, *2*), where Roche (*3*) had shown that strong tidal forces prevent the formation of any big satellite. Pan is located in the A ring's Encke Gap at 133,600 km from Saturn's center, and Atlas orbits at 137,700 km from Saturn's center, just outside the A ring. The Cassini spacecraft has recently resolved them both. Their shapes (Fig. 1) are close to oblate ellipsoids, with equatorial radii of 16.5 and ~19.5 km, and polar radii of ~10.5 km and 9 km for Pan and Atlas, respectively. These dimensions (*4*) are close to the moons' Hill radii (corresponding to the satellites' gravitational cross sections). More unexpectedly, both have a prominent equatorial ridge. These ridges are roughly symmetric about the bodies' equators and give them the appearance of a "flying saucer." Assuming that Pan and Atlas are rotating synchronously around Saturn (like the Moon around the Earth), consistent with Cassini images taken at several different times (*5*), Pan's ridge extends from –15° to +15° latitude (±~5°) and apparently entirely encircles the satellite. Atlas' ridge extends from –30° to +30° latitude (±10°) on the trailing side, whereas on the leading side the ridge is much less prominent, with a modest depression on the leading side near the equator (*4*) (Fig. 1C).

Recent work (*6*) has shown that a fast rotation may explain the diamond shape of the near-earth asteroid 1999 KW4 because of the balance of the centrifugal and gravity forces at the asteroid's equator. This mechanism seems, however, inadequate to explain the shapes of Pan and Atlas: Their rotation periods T (~14 hours) are much too long for centrifugal forces to balance surface gravity (which requires $T \sim 5$ hours). In addition, Saturn's tidal stress would elongate the moons in the radial direction (*3*, *7*) rather than create an equatorial ridge. Therefore, neither centrifugal nor tidal forces seem adequate to explain these ridges.

A number of circumstances led us to investigate a different scenario for the creation of the ridges: (i) Contrary to other resolved satellites, Pan and Atlas are embedded in Saturn's rings. (ii) The ridges are equatorial and precisely in the same plane as Saturn's rings. (iii) The vertical motion of Atlas (and perhaps Pan) through the rings is approximately equal to the

[1]Laboratoire AIM, Commissariat à l'Énergie Atomique (CEA)/Université Paris 7/CNRS, 91191 Gif-sur-Yvette Cedex, France. [2]Center for Radiophysics and Space Research, Cornell University, Ithaca, NY 14853, USA. [3]Cassini Imaging Central Laboratory for Operations, Space Science Institute, Boulder, CO 80301, USA.

*To whom correspondence should be addressed: charnoz@cea.fr

vertical extent of the ridges (*8*). (iv) The total volume of the ridges is ~10 to 25% of the bodies' volume (*4*). Therefore, we have explored the possibility that the ridges are made of ring particles accreted lately onto the surface of a preexisting moonlet embedded in the rings and initially free of any equatorial ridge. Then, a ridge would be simply an "equatorial ornament" accumulated onto the body's surface as a later stage in the formation process.

For a satellite with an orbital inclination i, semimajor axis a_s, and radius r, the latitudinal extension of the intersection of the ring plane with the body's surface is $L = \sin^{-1}(ia_s/r)$, assuming a vertical thickness h of the rings, much smaller than the satellite radius r, consistent with observations and models (*9*, *10*). Ring particles may collide between latitudes –L and +L. Using the average radii of Pan and Atlas (*4*) and their last published inclinations (0.001° and 0.003°, respectively) (*8*), we get, respectively, $L = \pm 12°$ and $L = \pm 28°$, in good agreement with observations. Atlas's inclination likely results from Prometheus's gravitational perturbations (*5*, *8*).

To simulate the fall of ring particles onto the surface of Pan and Atlas, a dynamical code (*11*) is used in which the orbits of Pan, Atlas, and 10^4 massless test particles are integrated, accounting for Saturn's J2 and J4 gravitational moments (*12*). The Pan and Atlas precursors have shapes similar to their actual ones but with equatorial radii smaller by ~10% to account for the initial absence of the ridges. Synchronous rotation is assumed. The test particles are initially gathered into a thin ring with vertical thickness $h = 250$ m, so that $h/r << 1$. Locations of impacts at the surface of the satellites are detected and located in the satellite's east longitude and latitude system (Fig. 2). Because of computer limitations, collisions between ring particles were ignored, but this approximation does not alter the latitudinal impact distribution as long as the ring system remains thin during the ridge accretion process, which is indeed the case throughout the simulation (SOM Text). Collisions would, however, modify the final longitudinal distribution, especially when the ridge accumulates on the satellite surface, so our simulation must be considered as a first-order model of the initial steps of ridge accretion.

The simulations show that impacts are concentrated near the equator (Fig. 2), with latitudinal ranges ±10° for Pan and ±30° for Atlas, in agreement with the observations (Fig. 1, A and B). Impacts on Atlas's surface near its leading (270° longitude) and trailing (90° longitude) points show a wider spread in latitude: This is a consequence of Atlas's much flatter shape compared with Pan. Unexpectedly, ring-particles' distribution of impacts appears to be strongly segregated: Particles with initial semimajor axes inward of their respective satellite ("inner particles") and those outward of the satellite ("outer particles") impact on different hemispheres (crosses and open diamonds in Fig. 2). For Pan, inner particles impact on the hemisphere facing Saturn and outer particles impact on the hemisphere opposite Saturn. For Atlas, the distributions are shifted by 90° compared with Pan: Inner particles collide on the leading hemisphere, and outer particles collide on the trailing hemisphere.

These different distributions are likely the consequences of the satellites' different eccentricities (3.5×10^{-5} and 1.2×10^{-3} for Pan and Atlas, respectively) (*8*). An orbital eccentricity e implies a radial excursion $\Delta r = \pm a_s e$. For Pan, $\Delta r = \pm 5$ km, which is much smaller than Pan's Hill radius of 19.5 km. Conversely, the Atlas radial excursion is ±165 km, much larger that its Hill radius of 23 km. Thus, different accretion processes must be considered for the two satellites: a low-velocity scenario for Pan and a high-velocity one for Atlas.

As a result of Pan's (almost) circular orbit, ring particles reach Pan's surface with low relative velocities. Therefore, classical celestial mechanics (*7*) tells us that particles must pass through Pan's L1 or L2 Lagrange points (the gates for entering the moon's Hill sphere). Inner particles flow through the L1 point (facing the sub-Saturn point), and outer particles flow through the L2 point (facing the anti-Saturn point). Once a particle penetrates Pan's Hill sphere, its trajectory has no space in which to randomize because the satellite fills almost all

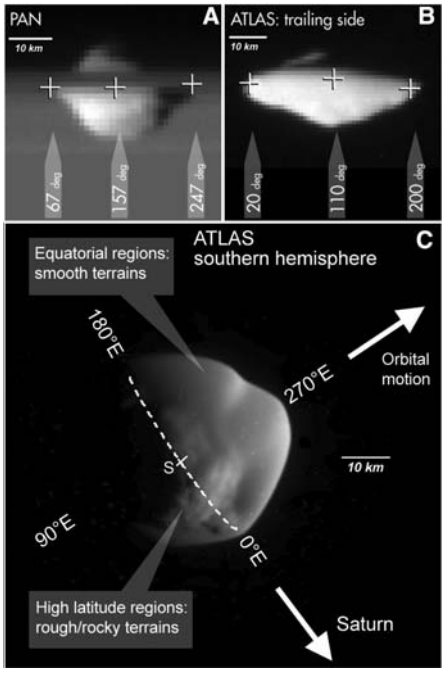


Fig. 1. Cassini narrow-angle camera (NAC) images in clear filter. (**A**) Pan in the Encke gap (1.3 km/pixel). Crosses are located on the body's equator with their corresponding longitude. (**B**) Atlas's trailing side with resolution 1.05 km/pixel. (**C**) High-resolution image of Atlas (320 m/pixel). The South Pole is designated by an "S"; the dashed line is the frontier between the leading and trailing sides. An east-longitude system is used, in which longitudes 0°, 90°, 180°, and 270° correspond to the sub-Saturn points, trailing points, anti-Saturn points, and leading points, respectively. White spots are cosmic rays.

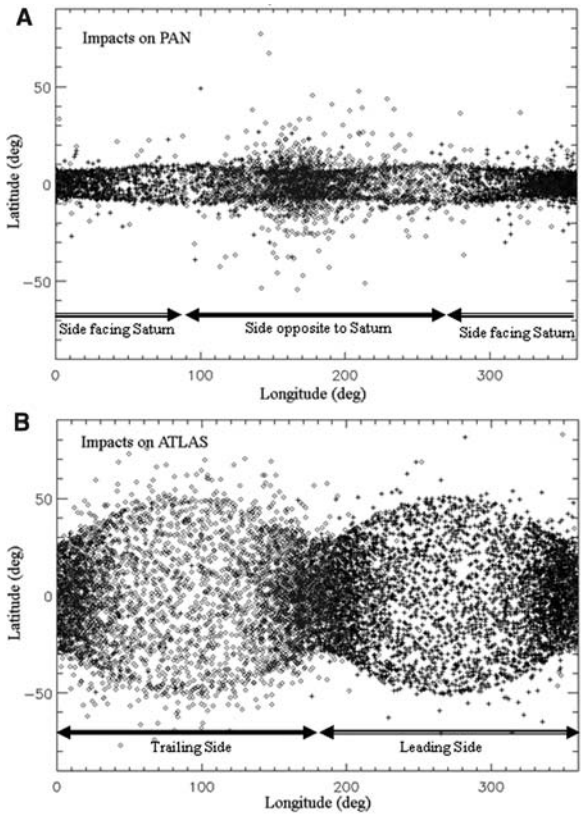


Fig. 2. Impact locations of particles at the surface of Pan (**A**) and Atlas (**B**). Open diamonds, particles initially exterior to the satellite's orbit; black crosses, particles initially interior to the satellite's orbit.

the space inside the Hill sphere (*4*). As a result, particle impacts Pan's surface almost immediately after passing the Lagrange point. This explains the segregation between the two hemispheres (Fig. 3), as well as the observation that the equatorial ridge encircles both hemispheres: Pan accretes material coming from both sides of its orbit.

Atlas, on the other hand, has a large eccentricity. Thus, in the local rotating frame, Atlas makes a clockwise elliptical epicycle with a 2:1 ratio (*7*), with its radial excursion Δr extending much beyond its Hill sphere. In the inner portion of the epicycle in which $r_s < a_s$, (where r_s is the instantaneous satellite's distance to Saturn), Atlas orbits at a higher velocity than the local Keplerian velocity, so inner particles are accreted onto Atlas's leading hemisphere. Conversely, on the outer portion of its epicycle ($r_s > a_s$), Atlas orbits at a lower velocity than the local Keplerian velocity, so outer particles are accreted on the trailing hemisphere only (Fig. 3B). This peculiar local dynamic may explain the differences observed between the trailing and leading sides of Atlas: It may have accreted material coming preferentially from outside its orbit rather than from inside, resulting in a less prominent bulge on the leading hemisphere (*4*).

Therefore, Pan and Atlas may have formed in two steps: an early stage in which a primordial body formed with roughly the current ellipsoidal shape but without an equatorial ridge, and a secondary stage in which the equatorial ridge is accreted from ring material. Consequently, differences between polar and equatorial terrains may be expected. This prediction has been confirmed by a June 13, 2007, high-resolution Cassini image (*4*) of Atlas's southern hemisphere (Fig. 1C). It shows two types of terrains: high-latitude regions (>40° south latitude) with a rough surface texture and equatorial regions (<30° south latitude) with a very smooth surface and no structure down to pixel resolution. A smooth surface is compatible with an accumulation of ring particles that are known to range from millimeters to meters, well below the camera resolution (*13*). Simulations also show that impacts on Pan and Atlas occur preferentially at sub- and anti-Saturn points (fig. S1), implying more elongated ridges in their radial direction. This seems marginally inconsistent with observations (*4*) (Fig. 1C). Our results apply to the early phase of ridge formation and, in reality, later evolutionary processes such as material redistribution, satellite reorientation, and the escape of material through the L1 and L2 Lagrange points will alter the final shape.

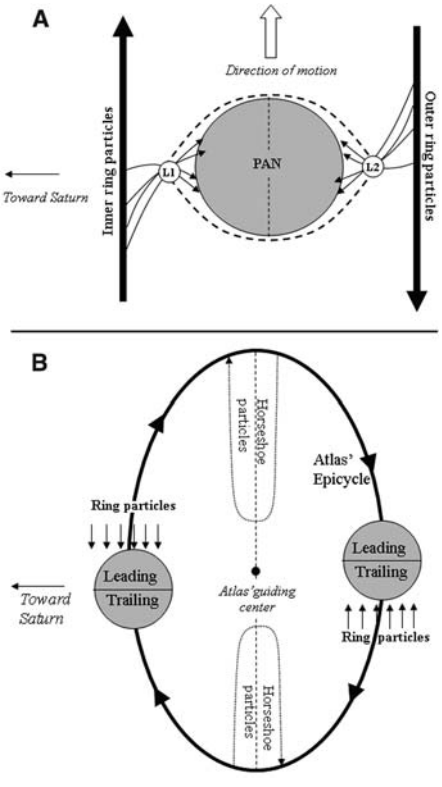


Fig. 3. (**A**) Sketch of the accretion process for Pan. Pan is on a quasicircular orbit. Arrows indicate trajectories of particles entering the Hill sphere of Pan. L1 and L2 stand for the two Lagrange points. (**B**) Sketch of the accretion process for Atlas. Relative sizes of Atlas and its epicycle are not to scale.

To have ridges confined to the equator, the incoming material must remain on very low inclination orbits ($<2 \times 10^{-3}$°) despite the satellite's gravitational stirring (*14*). Therefore, either (i) the ridge is accreted rapidly, as simulated here, or (ii) a dissipative environment is present to damp incoming particles' inclinations. A recent work (*4*) suggests that Pan, and maybe Atlas, opened a gap before completing accretion. So, ridge accretion may have happened after the start of the gap opening but before the complete emptying of the satellites' surroundings. This suggests that a transition phase may have existed: while a gap was already opened, the satellite could be fed by an accretion disk flowing from the gap's edges, like for giant-planet formation in the protoplanetary disk. If needed, dissipative collisions would also lower Pan's inclination. Conversely, Atlas's inclination, being gravitationally forced by Prometheus (*8*), remains nonzero

Voyager (*2*) and Cassini images (*15*) show the presence of tenuous material in the Pan and Atlas regions, and we examined whether this material could be accumulated onto the ridges. Our simulations show (fig. S2), on the one hand, that the material in the ringlet close to Pan is prevented from reaching Pan's surface because it is on horseshoe orbits and Pan is on a nearly circular orbit. On the other hand, because of Atlas's substantial inclination, the surrounding material is vertically stirred rapidly (fig. S3) and is prevented from accumulating specifically at Atlas's equator. So, a recent accretion of ridges from the material in the Encke gap or in the Atlas region seems unlikely, although some contribution cannot be completely discarded.

These results, together with other evidence collected by Cassini (*4*), suggest a relative chronology for the formation of Saturn's rings and the ridges of Pan and Atlas. The process by which Saturn's rings formed is still open to discussion (*16*, *17*). However, if a catastrophic breakup of a larger body happened, the cores of Pan and Atlas would have been among the largest fragments, orbiting initially on an inclined and eccentric orbit. Then, a two-step process may have occurred: Before the flattening of the rings, these shards accreted a shell of debris coming from the surrounding disk, giving them their low density and overall Roche lobe size (*4*). Once the debris disk flattened into a thin ring system due to dissipative collisions, ring material accreted at the satellites' equator through an accretion disk as outlined herein, forming today's ridges as observed by Cassini.

These ridges may also be of interest for planetary formation: They could be considered as "fossilized" accretion disks that once may have surrounded Pan and Atlas, like small-scale versions of the planetary subnebulas that once surrounded the giant planets. Such fossilized disks may result from two extreme characteristics of small satellites in planetary rings: (i) In rings, the ratio of the disk's thickness to the satellite's size is much smaller than unity, so that material flows toward the satellite's equator. (ii) In rings, it has been found that these satellites completely fill their Hill spheres (*4*) so that the trapped material is squeezed at the body's surface and accumulates immediately, forming today's equatorial ridges.

References and Notes

1. M. R. Showalter, *Nature* **351**, 709 (1991).
2. B. A. Smith *et al.*, *Science* **212**, 163 (1981).
3. E. Roche, *Mémoires de la Section des Sciences. Académie des Sciences et Lettres de Montpellier* **1**, 243 (1849).
4. C. C. Porco, P. C. Thomas, J. W. Weiss, D. C. Richardson, *Science* **318**, 1602 (2007).
5. C. C. Porco *et al.*, *Bull. Am. Astr. Soc.* **37**, 768 (2005).
6. S. J. Ostro *et al.*, *Science* **314**, 1276 (2006).
7. H. Poincaré, *Leçons de Mécanique Céleste* (Gauthier-Villars, Paris, 1905).
8. J. N. Spitale, R. A. Jacobson, C. C. Porco, W. M. Owen, *Astron. J.* **132**, 692 (2006).
9. A. Brahic, B. Sicardy, *Nature* **289**, 447 (1981).
10. J. A. Cowell, L. W. Esposito, M. Sremčević, *GeoRL* **33**, L07201 (2006).
11. S. Charnoz, P. Thébault, A. Brahic, *Astron. Astrophys.* **373**, 683 (2001).
12. J. K. Campbell, J. D. Anderson, *Astron. J.* **97**, 1485 (1989).
13. H. A. Zebker, E. A. Marouf, G. L. Tyler, *Icarus* **64**, 531 (1985).
14. R. Karjalainen, H. Salo, *Icarus* **172**, 328 (2004).
15. C. C. Porco *et al.*, *Science* **307**, 1226 (2005).
16. A. Harris, in *Planetary Rings*, R. Greenberg, A. Brahic, Eds. (Univ. Arizona Press, 1984), pp. 641–659.
17. M. S. Tiscareno *et al.*, *Nature* **440**, 648 (2006).
18. C.C.P. and P.C.T. acknowledge the support of NASA/JPL and the Cassini Project. We thank J. Colwell and H. Throop for careful reading, F. Durillon for graphics, and three anonymous referees.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1622/DC1
SOM Text
Figs. S1 to S3

1 August 2007; accepted 4 October 2007
10.1126/science.1148631

Stabilizing Isopeptide Bonds Revealed in Gram-Positive Bacterial Pilus Structure

Hae Joo Kang,[1,2] Fasséli Coulibaly,[1,2] Fiona Clow,[1,3] Thomas Proft,[1,3]* Edward N. Baker[1,2]*

Many bacterial pathogens have long, slender pili through which they adhere to host cells. The crystal structure of the major pilin subunit from the Gram-positive human pathogen *Streptococcus pyogenes* at 2.2 angstroms resolution reveals an extended structure comprising two all-β domains. The molecules associate in columns through the crystal, with each carboxyl terminus adjacent to a conserved lysine of the next molecule. This lysine forms the isopeptide bonds that link the subunits in native pili, validating the relevance of the crystal assembly. Each subunit contains two lysine-asparagine isopeptide bonds generated by an intramolecular reaction, and we find evidence for similar isopeptide bonds in other cell surface proteins of Gram-positive bacteria. The present structure explains the strength and stability of such Gram-positive pili and could facilitate vaccine development.

Bacterial pili are filamentous structures that extend from the bacterial cell surface and mediate host cell adhesion, bacterial motility, and other critical aspects of colonization. The pili of pathogenic bacteria are also major virulence factors and important vaccine candidates. The best-characterized are the type I and type IV pili of Gram-negative organisms, for which considerable structural information exists on subunit structure and assembly (*1*–*6*). These pili are long (1 to 4 μm), thin (5 to 8 nm), and flexible, but are nonetheless very strong and can withstand extreme physical stresses.

By contrast, the pili on Gram-positive bacteria have mostly gone unrecognized until recently, probably because they are extremely thin (2 to 3 nm) and hard to see. Unlike Gram-negative pili, whose subunits associate via noncovalent interactions, Gram-positive pili are assembled by bacterially encoded transpeptidase enzymes called sortases. These enzymes recognize specific sequence motifs in the pilin subunits, elongate the pilus oligomer by progressive addition of subunits joined by intermolecular isopeptide bonds, and then tether the entire assembly to the cell wall peptidoglycan (*7*–*9*). The pili thus consist of multiple, covalently bonded copies of a single backbone pilin, to which can be added a few accessory proteins.

Streptococcus pyogenes [group A *Streptococcus* (GAS)] infects the human throat and skin, causing common infections such as sore throat and tonsilitis, as well as severe invasive illnesses such as necrotizing fasciitis, rheumatic fever, and streptococcal toxic shock syndrome (*10*). Thin pili, ~2 nm wide and >1 μm long, have been revealed by electron microscopy (*11*) and were shown to be essential for adhesion to human tonsil and skin cells (*12*) as well as promising vaccine candidates against virulent GAS bacteria (*11*). The pilus-forming proteins are encoded in a small gene cluster within a pathogenicity island known as the FCT (fibronectin-binding, collagen-binding T antigen) region. In the *S. pyogenes* M1 strain SF370, *spy0128* encodes the backbone pilin, *spy0129* the sortase C1, and *spy0125* and *spy0130* two pilin-associated proteins (*11*). The backbone pilin subunits are Lancefield T antigens (*13*, *14*), named for their antigenicity and their extreme resistance to trypsin (T) digestion.

To understand pilus stability and assembly in Gram-positive organisms, we expressed the backbone pilin protein Spy0128 from an M1 strain of *S. pyogenes*. This 340-residue protein has a sortase recognition motif, Glu-Val-Pro-Thr-Gly, at residues 308 to 312. Constructs comprising residues 18 to 311 and 18 to 308 were prepared. We obtained excellent crystals for the latter (*15*) and solved its crystal structure at 2.2 Å resolution (R = 20.3%, R_{free} = 26.4%) (table S1).

The Spy0128 monomer has an elongated two-domain structure, with length 98 Å and width 20 to 30 Å (Fig. 1A). Both domains have irregular all-β structures that are modified variants of the immunoglobulin fold (Fig. 1B). The N-terminal domain, residues 18 to 171, forms a β sandwich in which the strands in one β sheet (green in Fig. 1) are progressively extended such that the upper portion of this β sheet, at the top of the domain as shown in Fig. 1, is relatively exposed. The C-terminal domain,

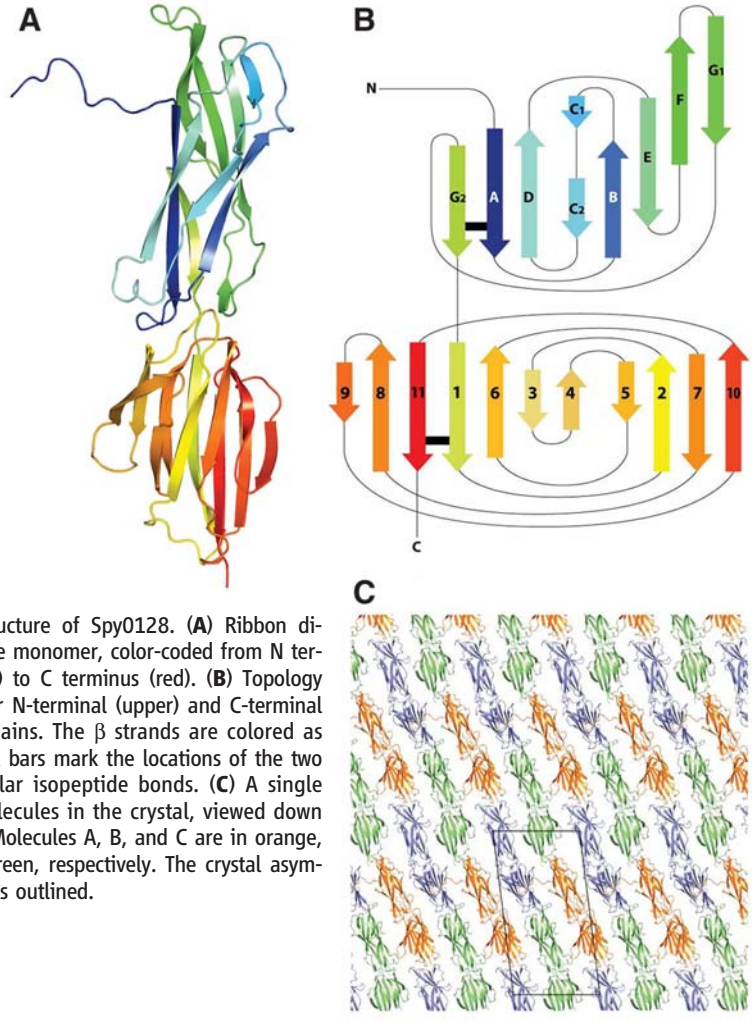


Fig. 1. Structure of Spy0128. (**A**) Ribbon diagram of the monomer, color-coded from N terminus (blue) to C terminus (red). (**B**) Topology diagrams for N-terminal (upper) and C-terminal (lower) domains. The β strands are colored as in (A). Black bars mark the locations of the two intramolecular isopeptide bonds. (**C**) A single layer of molecules in the crystal, viewed down the *b* axis. Molecules A, B, and C are in orange, blue, and green, respectively. The crystal asymmetric unit is outlined.

[1]Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland 1010, New Zealand. [2]School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand. [3]School of Medical Sciences, University of Auckland, Auckland 1023, New Zealand.

*To whom correspondence should be addressed. E-mail: ted.baker@auckland.ac.nz (E.N.B.); t.proft@auckland.ac.nz (T.P.)

residues 173 to 307, comprises 11 β strands. Its core is a β sandwich in which a five-stranded β sheet packs against a four-stranded β sheet. A prominent β ribbon (strands β3 and β4) extends the first sheet to seven strands and provides a wide loop at the base of the domain. Overall, the domain is wedge-shaped with a broad base and a narrower top where it joins to the N domain. The two domains are intimately associated, with only one residue, Ser^{172}, separating the final β strand of the N domain from the first of the C domain. The interface between domains is mostly hydrophobic and buries ~1200 $Å^2$ of surface area.

The crystal asymmetric unit contains three independent Spy0128 molecules that generate columns of molecules extending through the crystal (Fig. 1C). This arrangement, found also in another crystal form (*15*), provides a compelling model for the assembly of GAS pili. Successive molecules stack head-to-tail, related by an approximate 3_1 helical screw along their long axis. Each interface, between the N domain of one molecule and the C domain of the next (Fig. 2), buries ~850 $Å^2$ of solvent-accessible surface with a shape complementarity of 0.72, comparable with other protein oligomerization interfaces (*16*). There is very little lateral interaction between columns of molecules in the crystal.

The head-to-tail packing means that Phe^{307}, which closely precedes the sortase recognition motif in Spy0128, packs against the exposed face of the N-domain β sheet (Fig. 2B). Sortase action cleaves the Thr^{311}-Gly^{312} bond, after which isopeptide bond formation between the new C terminus and a Lys residue covalently links adjacent pilin subunits (*7*, *8*). Five invariant lysines (Fig. 2C) are potential candidates for this intermolecular linkage. Of these, only Lys^{161}, near the top of the N domain and 11 to 13 Å below Phe^{307} of the next molecule in the column, is a viable candidate for generating an elongated pilus. We used mass spectrometry of pilus fractions from *S. pyogenes* to show that Lys^{161} is indeed the essential lysine involved in oligomerization (fig. S3). This finding strongly supports the biological relevance of the assembly seen in all crystal forms. Residues 308 to 311 would continue below Phe^{307}, packing against a highly sequence-conserved region of the β sheet (Fig. 2B) and allowing isopeptide bond formation between the Thr^{311} carboxyl and Lys^{161} Nζ of the next molecule.

Intermolecular isopeptide bonds are known in other contexts besides the sortase-generated isopeptide bonds of Gram-positive pili. In ubiquitination, specific lysine residues of a target protein are covalently linked by ubiquitin ligases to the terminal carboxylate of ubiquitin (*17*). In transglutamination, enzyme-catalyzed isopeptide bond formation occurs between Gln and Lys side chains (*18*), as in the cross-linking of fibrin subunits, catalyzed by factor XIII (*19*). A rare example of self-generated isopeptide bonds between Asn and Lys residues occurs in the bacteriophage HK97, where capsid subunits are covalently cross-linked to form interlocked circular rings that give extraordinary stability (*20*, *21*). No examples of intramolecular isopeptide bonds have been reported, however.

In this context we were surprised to observe two intramolecular isopeptide bonds within the pilin subunit, one in each domain. Formed by covalent bonding between lysine and asparagine side chains (Lys^{36}-Asn^{168} in the N domain; Lys^{179}-Asn^{303} in the C domain), these are each indicated by continuous electron density extending through the lysine ε-amino group into the δ-carboxyamide group of asparagine (Fig. 3A). Mass spectrometry provided independent confirmation (*15*). The protein molecular mass was consistent with the loss of two NH_3 units through isopeptide bond formation (table S2), and proteolytic digestion and peptide mapping gave cleavage products containing nonconsecutive sequences (figs. S1 and S2). These mapped to peptides surrounding both isopeptide bonds.

These bonds appear, as in HK97, to be self-generated. An essential Glu residue is associated with each bond, forming hydrogen bonds to the isopeptide C=O and NH groups (Fig. 3A). The hydrogen bonding implies that both glutamic acids, Glu^{117} and Glu^{258}, are protonated. In each case, the Lys, Asn, and Glu residues are surrounded by a cluster of aromatic residues (Fig. 3C), which would favor elevation of the pK_a of the glutamic acid and reduction of the pK_a of the lysine ε-amino group. In the N domain, the isopeptide moiety sits over the aromatic ring plane of Phe^{52}, and Glu^{117} is surrounded by Phe^{54}, Tyr^{128}, and Phe^{166}. Similar roles are played by Phe^{192}, Phe^{194}, Tyr^{261}, and Phe^{301} for the C-terminal isopeptide. A plausible mechanism for isopeptide bond formation, first suggested for HK97 (*21*), is that the protonated Glu polarizes the C=O bond of the Asn side chain, inducing positive charge on Cγ. Nucleophilic attack on Cγ by the unprotonated Lys ε-amino group then generates the isopeptide bond.

We tested the importance of the Glu residues for isopeptide formation by mutating Glu^{117} and Glu^{258} to alanine, creating proteins E117A and E258A. Mass spectrometry showed the loss of one isopeptide bond from each mutant (table S2), and crystallographic analysis of E117A confirmed that the N-domain isopeptide was not formed when Glu^{117} was mutated (fig. S4). Both mutants also showed greatly increased susceptibility to proteolysis, indicating the stabilizing effect of these cross-links (fig. S5).

Sequence comparisons (fig. S6) suggested that the isopeptide bonds may be a conserved feature of the pili of all GAS. Despite low over-

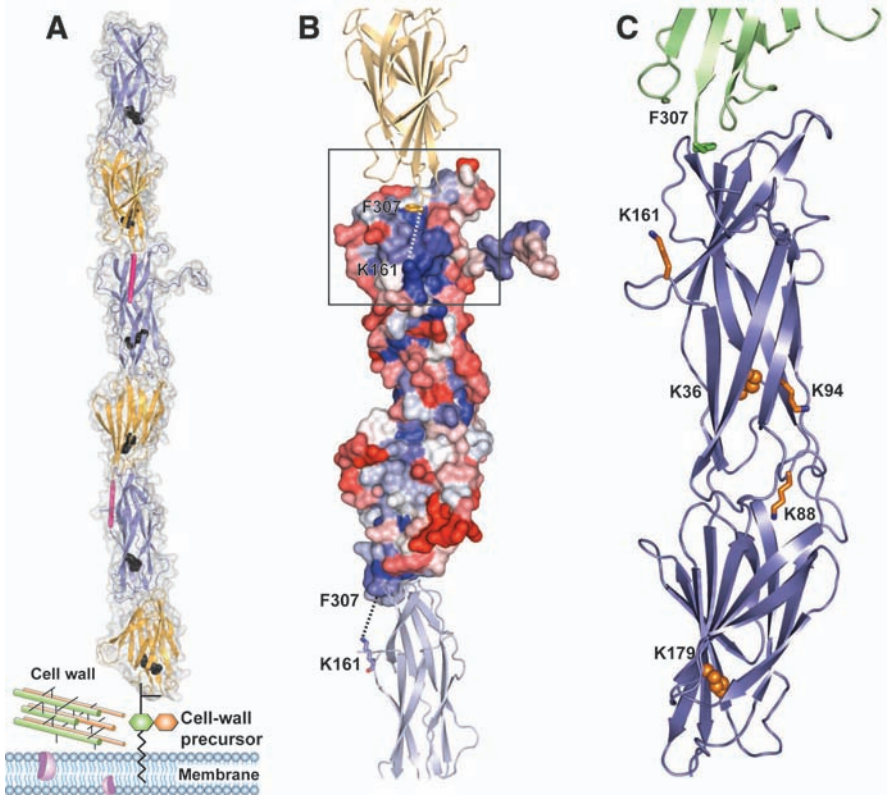


Fig. 2. Assembly of *S. pyogenes* pili. (**A**) Column of Spy0128 molecules in which the N domain (blue) of one molecule packs against the C domain (gold) of the next. Black bars indicate the intramolecular isopeptide bonds and red bars the proposed intermolecular links. (**B**) Pilus assembly in which one molecule is shown in surface representation, colored by sequence conservation across GAS major pilins (table S2) from dark blue (highly conserved) to dark red (highly variable). A broken line shows the location of the sortase-mediated intermolecular isopeptide bond, joining the C terminus to the invariant Lys^{161}. F, Phe; K, Lys. (**C**) Location of all conserved lysines in Spy0128.

all sequence identity in Spy0128 alleles, the Lys, Asn, and Glu residues of the isopeptide bonds are strictly conserved, as are five of the eight aromatic residues surrounding them. The other aromatics are replaced only by hydrophobic residues. The isopeptide bonds are strategically located in each domain (just before the interdomain connection and the sortase recognition motif, respectively), tying together the first and last β strands (Fig. 1B). Sequence similarities with the major pilins from other Gram-positive bacteria are too low to determine whether isopeptide bonds are a common feature, but a conserved Asn precedes the sortase motif by 5 to 8 residues in all sequences we have examined (fig. S7), and conserved Lys and Glu residues can also be traced.

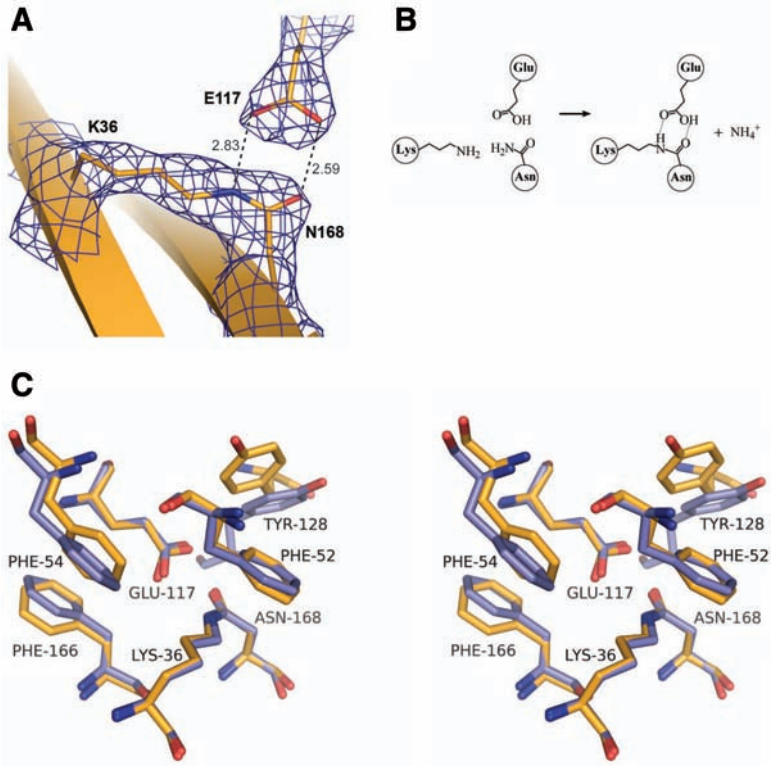


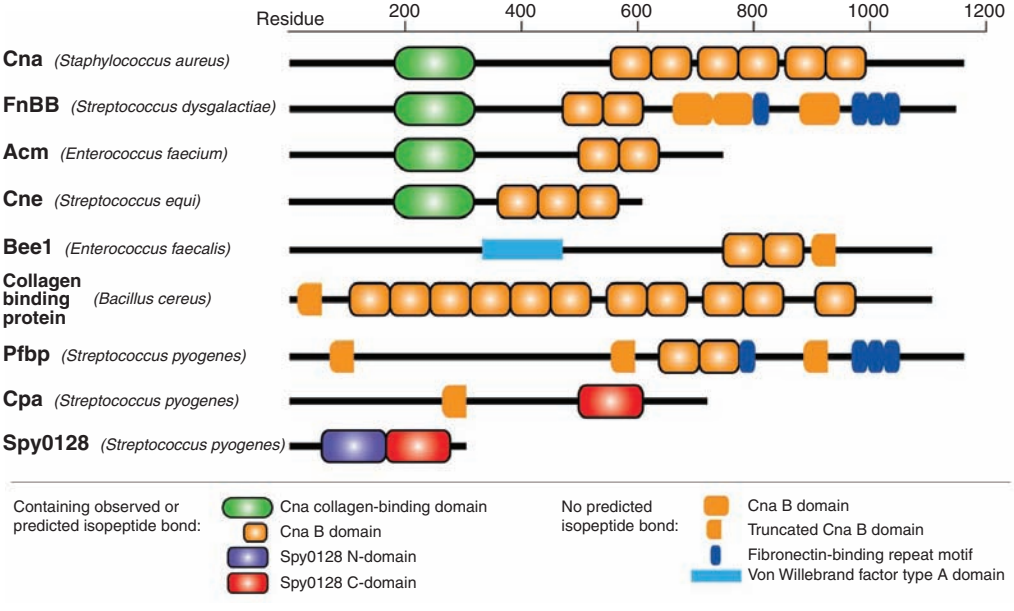
Fig. 3. Isopeptide bond formation in Spy0128. (**A**) The Lys-Asn (K-N) isopeptide bond in the N domain, shown in stick mode, in its $F_{obs} - F_{calc}$ omit density, contoured at 1.5σ. Hydrogen bonds with the catalytic Glu^{117} (E117) are shown with broken lines, with distances in Å. (**B**) The chemical reaction for isopeptide bond formation. (**C**) Stereo view of a superposition of the N-terminal (blue) and C-terminal (orange) isopeptide bonds, with the conserved aromatic residues surrounding them.

We also found evidence for intramolecular isopeptide bonds in other cell surface proteins. The C-terminal domain of the pilin-associated Cpa (GAS collagen-binding protein), encoded by *spy0125*, is homologous with the C domain of Spy0128, with residues involved in the C-terminal isopeptide bond (Lys, Asn, Glu, and three Phe) invariant across all 14 Cpa sequences in the current sequence database (fig. S8). Examination of the recently released structure of a minor pilin, GBS52 from *Streptococcus agalactiae* (*22*), reveals an unrecognized Lys-Asn isopeptide bond like those in Spy0128 (fig. S9). We then searched the Protein Data Bank, using a Lys-Asn-Glu/Asp structural template, and identified the collagen-binding adhesin Cna from *Staphylococcus aureus* (*23*, *24*) as also having previously unrecognized isopeptide bonds in its A and B domains (figs. S9 and S10). Further sequence searches showed many instances of these domains containing predicted isopeptide bond–forming residues in the same locations (Fig. 4), all from Gram-positive organisms and all (where functionally characterized) cell surface adhesion proteins.

The isopeptide bonds we have found in GAS pili and other Gram-positive adhesins provide a striking parallel with the disulfide bridges found in Gram-negative pilins and adhesins (*1*, *5*), which are important for pilus assembly and substrate binding. We hypothesize that in Gram-positive organisms, which lack the disulfide bond formation machinery of Gram-negative bacteria, intramolecular isopeptide bonds may provide an alternative mode of stabilization for cell surface proteins involved in host pathogenesis.

Our results provide a model for the assembly of *S. pyogenes* pili, in which self-generated intramolecular isopeptide bonds complement the sortase-catalyzed intermolecular bonds. The long, thin GAS pili are only ~2 nm (one molecule) thick (*11*) but typically >100 molecules long, and we infer that these bonds play a crit-

Fig. 4. Domain organization of proteins containing observed or predicted isopeptide bonds. Only proteins with known functions are shown. Domains with the conserved Lys-Asn-Glu of Spy0128 and CnaB or Lys-Asn-Asp of CnaA are highlighted by black outlines.



ical role in maintaining pilus integrity in the face of severe mechanical and chemical stress while bound to host cells. GAS pili show considerable antigenic variation, indicating an important role in virulence, and the pilin subunits are T antigens that are used for serotyping (*13*). The presence of several conserved regions on a highly variable background (Fig. 2B) suggests that the structure could help provide an effective pilus-based vaccine against GAS.

References and Notes

1. D. Choudhury *et al.*, *Science* **285**, 1061 (1999).
2. E. Hahn *et al.*, *J. Mol. Biol.* **323**, 845 (2002).
3. H. Remaut, G. Waksman, *Curr. Opin. Struct. Biol.* **14**, 161 (2004).
4. H. E. Parge *et al.*, *Nature* **378**, 32 (1995).
5. L. Craig, M. E. Pique, J. A. Tainer, *Nat. Rev. Microbiol.* **2**, 363 (2004).
6. L. Craig *et al.*, *Mol. Cell* **23**, 651 (2006).
7. H. Ton-That, O. Schneewind, *Mol. Microbiol.* **50**, 1429 (2003).
8. J. L. Telford, M. A. Barocchi, I. Margarit, R. Rappuoli, G. Grandi, *Nat. Rev. Microbiol.* **4**, 509 (2006).
9. H. Ton-That, L. A. Marraffini, O. Schneewind, *Biochim. Biophys. Acta* **1694**, 269 (2004).
10. M. W. Cunningham, *Clin. Microbiol. Rev.* **13**, 470 (2000).
11. M. Mora *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15641 (2005).
12. E. L. Abbot *et al.*, *Cell. Microbiol.* **9**, 1822 (2007).
13. R. C. Lancefield, *J. Exp. Med.* **71**, 521 (1940).
14. R. C. Lancefield, V. P. Dole, *J. Exp. Med.* **84**, 449 (1946).
15. See supporting material on *Science* Online.
16. M. C. Lawrence, P. M. Colman, *J. Mol. Biol.* **234**, 946 (1993).
17. C. M. Pickart, *Annu. Rev. Biochem.* **70**, 503 (2001).
18. C. S. Greenberg, P. J. Birckbichler, R. H. Rice, *FASEB J.* **5**, 3071 (1991).
19. J. W. Weisel, *Adv. Protein Chem.* **70**, 247 (2005).
20. W. R. Wikoff *et al.*, *Science* **289**, 2129 (2000).
21. C. Helgstrand *et al.*, *J. Mol. Biol.* **334**, 885 (2003).
22. V. Krishnan *et al.*, *Structure* **15**, 893 (2007).
23. C. C. S. Deivanayagam *et al.*, *Structure* **8**, 67 (2000).
24. Y. Zong *et al.*, *EMBO J.* **24**, 4224 (2005).

25. Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank with accession number 3B2M. We thank T. Caradoc-Davies for help with refinement and M. Middleditch for help with mass spectrometry. Supported by the Health Research Council (HRC) of New Zealand, the Foundation for Research, Science and Technology (FRST) of New Zealand, and the Maurice Wilkins Centre for Molecular Biodiscovery. H.J.K. received an FRST Bright Futures Scholarship, and T.P. is an HRC Hercus Fellow.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/PAGE/DC1
Materials and Methods
SOM Text
Figs. S1 to S13
Tables S1 to S4
References

29 May 2007; accepted 23 October 2007
10.1126/science.1145806

DUBA: A Deubiquitinase That Regulates Type I Interferon Production

Nobuhiko Kayagaki,[1] Qui Phung,[2] Salina Chan,[1] Ruchir Chaudhari,[1] Casey Quan,[1] Karen M. O'Rourke,[1] Michael Eby,[1] Eric Pietras,[3] Genhong Cheng,[3] J. Fernando Bazan,[4] Zemin Zhang,[5] David Arnott,[2] Vishva M. Dixit[1]*

Production of type I interferon (IFN-I) is a critical host defense triggered by pattern-recognition receptors (PRRs) of the innate immune system. Deubiquitinating enzyme A (DUBA), an ovarian tumor domain-containing deubiquitinating enzyme, was discovered in a small interfering RNA–based screen as a regulator of IFN-I production. Reduction of DUBA augmented the PRR-induced IFN-I response, whereas ectopic expression of DUBA had the converse effect. DUBA bound tumor necrosis factor receptor–associated factor 3 (TRAF3), an adaptor protein essential for the IFN-I response. TRAF3 is an E3 ubiquitin ligase that preferentially assembled lysine-63–linked polyubiquitin chains. DUBA selectively cleaved the lysine-63–linked polyubiquitin chains on TRAF3, resulting in its dissociation from the downstream signaling complex containing TANK-binding kinase 1. A discrete ubiquitin interaction motif within DUBA was required for efficient deubiquitination of TRAF3 and optimal suppression of IFN-I. Our data identify DUBA as a negative regulator of innate immune responses.

Innate immune responses are initiated when host cellular PRRs encounter pathogen-associated molecular patterns (PAMPs) (*1*). Double- and single-stranded RNAs are virus-derived PAMPs that trigger the intracellular PRRs Toll-like receptor 3 (TLR3), retinoic acid–inducible protein I (RIG-I), and melanoma differentiation–associated gene 5 (MDA5) (*2*–*4*). Activation of these intracellular sensors leads to the recruitment of adaptor proteins for interferon-α (IFN-α) and IFN-β production. Toll-interleukin 1 receptor domain–containing adaptor inducing IFN-β (TRIF) interacts with TLR3, whereas IFN-β promoter stimulator 1 [(IPS-1), also called Cardif, MAVS, and VISA] is recruited by RIG-I and MDA5. These adaptors mediate the assembly of a signaling complex composed of the ubiquitin ligase TRAF3 and the kinases TANK-binding kinase 1 (TBK1) and inhibitor of nuclear factor κB kinase ε [(IKKε), also called IKK*i* (*1*, *5*–*7*)]. This complex activates the downstream transcription factors, IFN regulatory factors 3 and 7 (IRF3 and IRF7), to switch on IFN-I expression, which is an essential aspect

[1]Department of Physiological Chemistry, Genentech, South San Francisco, CA 94080, USA. [2]Department of Protein Chemistry, Genentech, South San Francisco, CA 94080, USA. [3]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095, USA. [4]Department of Protein Engineering, Genentech, South San Francisco, CA 94080, USA. [5]Department of Bioinformatics, Genentech, South San Francisco, CA 94080, USA.

*To whom correspondence should be addressed. E-mail: dixit@gene.com

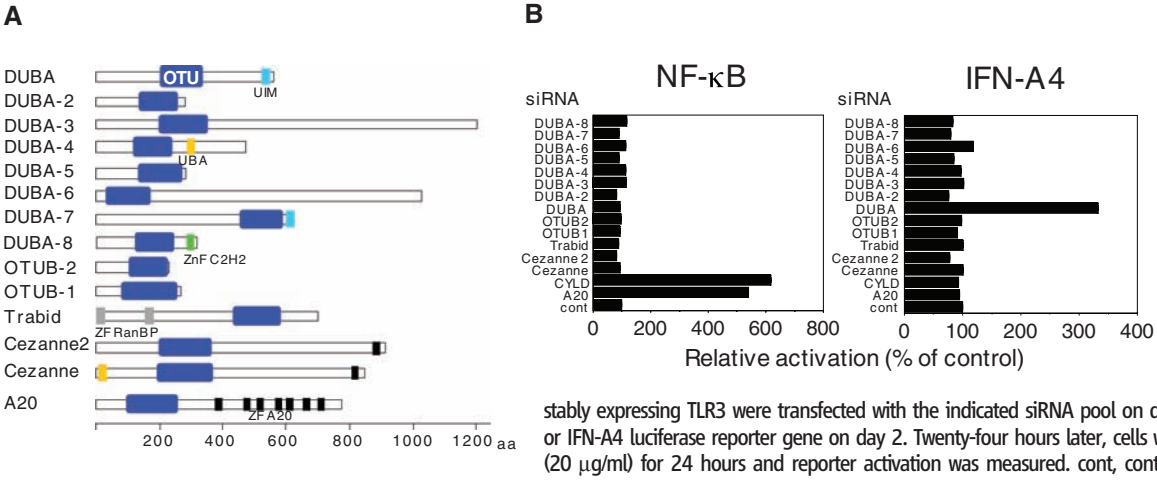


Fig. 1. An siRNA-based screen for OTU DUB family members. (**A**) Schematic of OTU family members. Gene accession numbers are listed in table S1. ZnF C2H2, zinc finger domain (C2H2-type); ZF RanBP, zinc finger domain (Ran-binding protein and others); ZF A20, zinc finger domain (A20-like); aa, amino acids. (**B**) HEK293 cells stably expressing TLR3 were transfected with the indicated siRNA pool on day 0 and then with a NF-κB or IFN-A4 luciferase reporter gene on day 2. Twenty-four hours later, cells were stimulated with poly(I:C) (20 μg/ml) for 24 hours and reporter activation was measured. cont, control.

of the innate immune response to invading viruses (*1*, *5*, *8*–*10*).

Ubiquitin is a protein found in all eukaryotic cells that gets covalently linked to various proteins by ubiquitin ligase enzymes (*11*). Lysines within ubiquitin can be modified so that polyubiquitin chains are formed; chains formed through linkages at Lys^{48} often mark target proteins for proteasome-mediated proteolytic destruction, whereas those linked through Lys^{63} can direct other outcomes. These chains have been implicated in the activation of signaling pathways that affect DNA repair or the transcription factor nuclear factor κB (NF-κB) (*11*, *12*). Deubiquitinating enzymes (DUBs) are proteases that specifically cleave ubiquitin linkages, negating the action of ubiquitin ligases (*13*). The DUB A20, for example, acts as a negative regulator of the classical NF-κB activation pathway (*14*). A20 belongs to a subfamily of 14 DUBs characterized by an ovarian tumor (OTU) domain (Fig. 1A). The physiological function of only a few members is known.

To identify potential roles of OTU subfamily DUBs, we used small interfering RNA oligonucleotides (siRNA oligos) to deplete various DUBs in TLR3-expressing human embryonic kidney (HEK) 293 cells. The impact on NF-κB activation or IFN-I production in response to the TLR3 ligand polyinosine:polycytidine acid [poly(I:C)] (which mimics viral RNA) was determined by luciferase reporter assay. Decreased A20 expression resulted in an increase in TLR3-induced NF-κB–dependent gene transcription (Fig. 1B). Activated TLR3 simultaneously promotes an IFN-I response, but this outcome is executed by the transcription factors IRF3 and IRF7 (*8*, *10*). The impact of A20 was limited to NF-κB activation because A20 reduction had a negligible effect on TLR3-induced activation of an IFN-A4 promoter (Fig. 1B). CYLD, a tumor suppressor belonging to the ubiquitin-specific protease subfamily of DUBs, is reported to be a negative regulator of NF-κB signaling (*12*, *13*, *15*, *16*). Depletion of CYLD resulted in increased TLR3-induced NF-κB–dependent gene transcription but did not increase activation of an IFN-A4 promoter (Fig. 1B). Similar to the classical NF-κB signaling

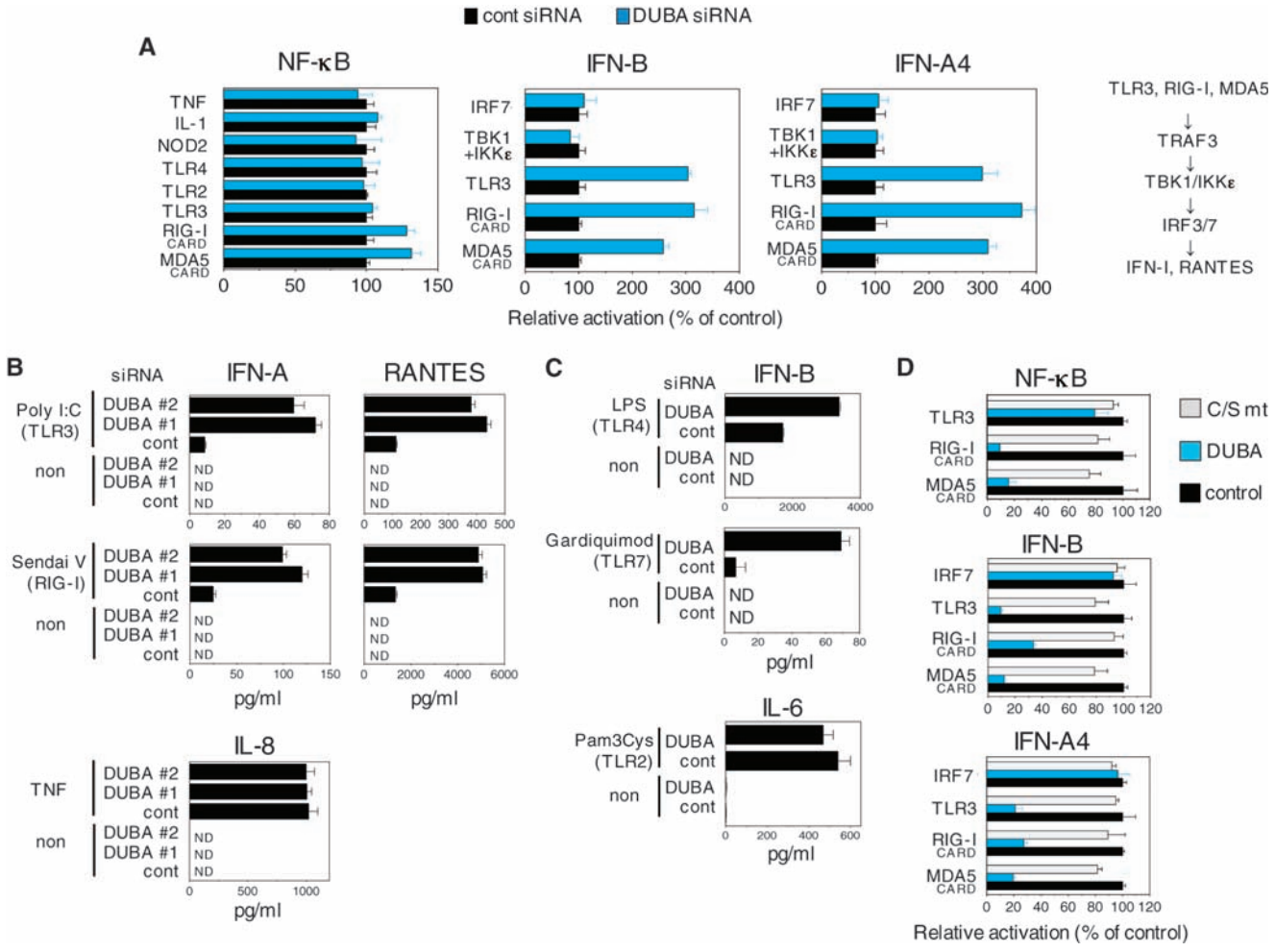


Fig. 2. Regulation of RIG-I and MDA5 signaling by DUBA. (**A**) HEK293 cells transfected with a DUBA siRNA (#1 in table S1) or a control siRNA were assayed for NF-κB activation or IFN-I (IFN-B and IFN-A4 promoter) activity. HEK293 cells were transfected with control or DUBA siRNA (20 nM) on day 0 and then with a NF-κB, IFN-B, or IFN-A4 luciferase reporter gene, together with indicated activators, on day 2. After 24 hours, ligands Pam3CSK4 (10 μg/ml), poly(I:C) (20 μg/ml), LPS (10 μg/ml), muramyldipeptide (10 μg/ml), IL-1 (50 ng/ml), or TNF-α (50 ng/ml) were added to the culture for 24 hours, and then reporter gene activation was measured. Data represent the mean ± SD of triplicate samples. (**B**) Parental HEK293 cells (middle and bottom panels) or those stably expressing TLR3 (top panels) were transfected with control or DUBA siRNAs. After 72 hours, cells were treated with poly(I:C) (20 μg/ml), Sendai virus [20 units (U)/ml], or TNF-α (50 ng/ml). Secreted IL-8, IFN-A, and RANTES were measured 24 hours later. Data represent the mean ± SD of triplicate samples. ND, not detectable. (**C**) DUBA function in macrophages. The macrophage cell line RAW264.3 was transfected with either a control siRNA or mouse DUBA siRNA and, 3 days later, cells were stimulated with the TLR ligands indicated. Cytokines secreted into the supernatant over the next 24 hours were measured. Data represent the mean ± SD of triplicate samples. (**D**) Effects of overexpressed DUBA. HEK293 cells were transfected with the indicated reporter gene and activators, together with either wild-type DUBA or the predicted catalytic active site mutant DUBA C224S (C/S mt). NF-κB or IFN-I reporter activity was measured after 36 hours. Data represent the mean ± SD of triplicate samples.

cascade, the IFN-I response is thought to require ubiquitination (*17*). We speculated that another DUB might negatively regulate IFN-I expression. Depletion of OTU subfamily member DUBA augmented TLR3-induced activation of an IFN-A4 promoter without altering activation of a NF-κB–dependent reporter (Fig. 1B). DUBA is predicted to be 571 amino acids long, and the catalytic residues (Asp^{221}, Cys^{224}, and His^{334}) essential for cysteine protease activity are conserved (fig. S1A). Northern blot analysis identified DUBA mRNA transcripts in various organs, including the liver and placenta, as well as in peripheral blood leukocytes (fig. S1B).

Two individual DUBA siRNA oligos increased TLR3-induced activation of an IFN-A4 promoter (fig. S2A), and depletion of endogenous DUBA protein was confirmed by Western blotting (fig. S2B). In the absence of TLR3 stimulation by poly(I:C), DUBA depletion did not influence responses to basal NF-κB or IFN-A4 (fig. S2A). Only double-stranded DUBA siRNA oligos increased IFN-A4 promoter activity (fig. S2C), thus excluding nonspecific engagement of the IFN-I and NF-κB pathways by the siRNAs (*18*).

RIG-I and MDA5 are cytosolic DExD/H-box RNA helicases that have a caspase activation and recruitment domain (CARD). These helicases detect invading RNA viruses and engage the IFN-I response through the adaptor protein IPS-1 (*1*, *4*, *5*). We expressed the CARD domains of RIG-I and MDA5 (RIG-I_{CARD} and $MDA5_{CARD}$) to limit possible nonspecific engagement of the helicase domains by the siRNA oligos. Depletion of DUBA increased activation of IFN-B and IFN-A4 promoters in response to RIG-I_{CARD} and $MDA5_{CARD}$ expression, and the magnitude of the effect was similar to that seen when poly(I:C) was used to stimulate TLR3 (Fig. 2A). The increased transcriptional response to poly(I:C) after DUBA depletion was concomitant with increased secretion of the cytokines IFN-A and RANTES, both of which are under the control of IRF3 and IRF7 (*8*, *10*) (Fig. 2B). Although depletion of DUBA led to a slight increase in the transcription of a NF-κB–dependent reporter gene in response to RIG-I_{CARD} and $MDA5_{CARD}$ overexpression, it did not alter NF-κB activation by tumor necrosis factor (TNF), interleukin-1 (IL-1), nucleotide-binding oligomerization domain 2 (NOD2), or TLRs 2 to 4 (Fig. 2A), nor did depletion of DUBA alter NF-κB–dependent TNF-induced secretion of IL-8 (Fig. 2B). Thus,

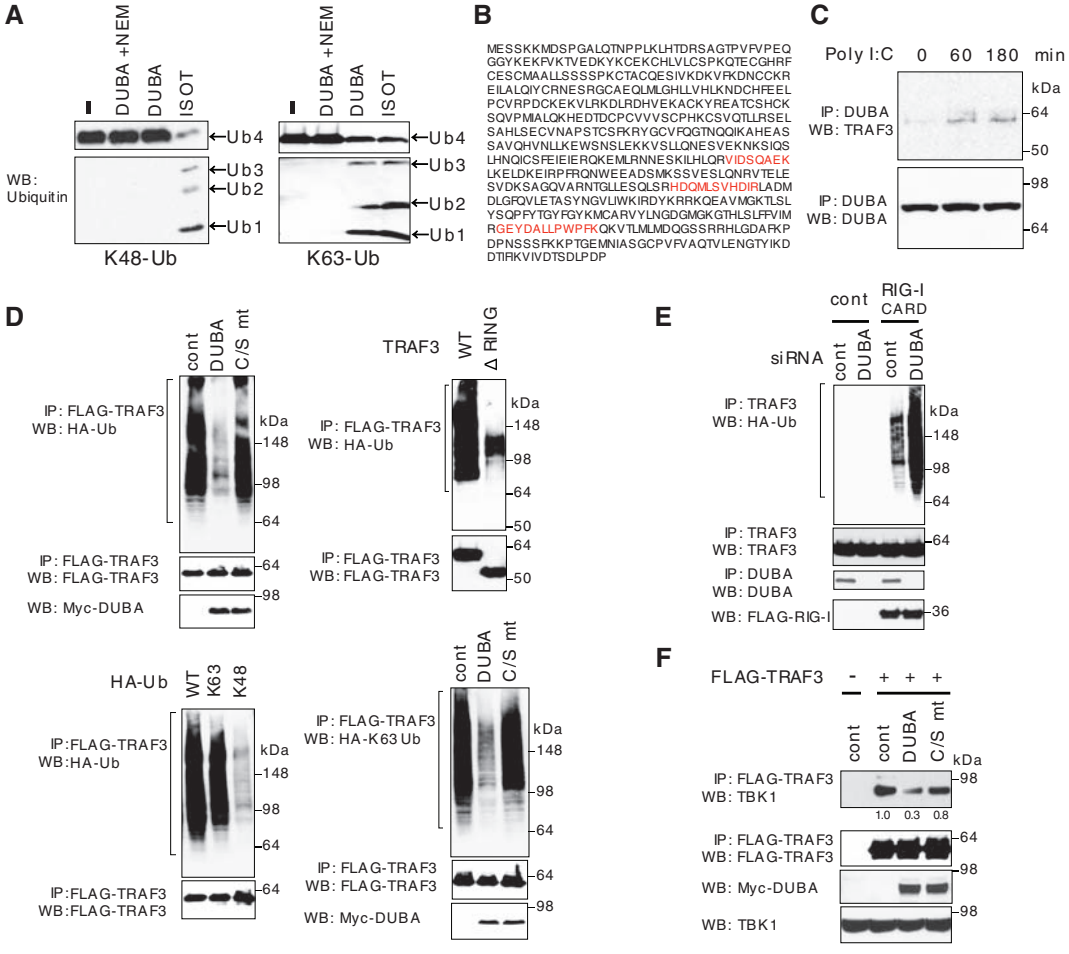


Fig. 3. Interaction of DUBA with TRAF3 and deubiquitination of TRAF3. (**A**) Ubiquitin isopeptidase activity of DUBA. Recombinant DUBA (1 μg) purified from *E. coli* or IsoT (positive control) was incubated with Lys^{48} (K48)–linked or Lys^{63} (K63)–linked tetra-ubiquitin chains (0.5 μg) at 37°C for 16 hours. Reactions were analyzed by Western blotting (WB) with ubiquitin antibody. Cysteine protease activity was blocked with 20 μM NEM in the reactions indicated. The top and bottom lanes in each panel are from different film exposure time points of the same membrane. (**B**) Sequence of TRAF3 and the matching endogenous peptides (highlighted in red) identified by mass spectrometry. The purified DUBA complex was analyzed by means of one-dimensional gel electrophoresis in combination with the nano liquid-chromatography tandem mass spectrometry. (**C**) Interaction of endogenous DUBA and TRAF3. HEK293 cells stably expressing TLR3 were stimulated with 20 μg/ml poly(I:C), and lysates were prepared at the time points indicated. Immunoprecipitation (IP) was performed with a DUBA antibody, and coprecipitating endogenous TRAF3 was detected by Western blotting. (**D**) Effects of exogenous DUBA on transfected TRAF3. HEK293 cells were cotransfected with FLAG-TRAF3, HA-tagged ubiquitin (HA-Ub), and with either empty vector, Myc-tagged wild-type (WT) DUBA, or DUBA C/S mt. Twenty-four hours later, cell lysate was subjected to heat denaturing in 1% SDS, TRAF3 was immunoprecipitated with FLAG antibody, and Western blotting was done with HA or FLAG antibodies. Wild-type TRAF3 or ΔRING TRAF3 mutant was used in the top right panel. Wild-type ubiquitin or mutants retaining only a single lysine (K48 or K63) were used in the bottom panels. 1% of input lysate was subjected to Western blotting with Myc antibody. (**E**) Ubiquitination of endogenous TRAF3. HEK293 cells were transfected with 20 nM control or DUBA #1 siRNA. Thirty-six hours later, cells were cotransfected with control or FLAG-tagged RIG-I_{CARD} and HA-Ub and were incubated for 24 hours. After heat denaturing in 1% SDS, endogenous TRAF3 was immunoprecipitated with anti-TRAF3 and subjected to immunoblotting with anti-HA or anti-TRAF3. 1% of input lysate was subjected to Western blotting with FLAG antibody. 50% of input lysate was subjected to immunoblotting for DUBA, as in fig. S2B. (**F**) DUBA dissociates TRAF3 from TBK1. HEK293 cells were transfected with FLAG-TRAF3 together with control plasmid, Myc-tagged DUBA, or DUBA C/S mt. After 48 hours, the TRAF3 complex was immunoprecipitated with anti-FLAG beads and subjected to immunoblotting with anti-TBK1 and anti-FLAG. 1% of input lysate was subjected to Western blotting with Myc and TBK1 antibodies. TBK1 was quantified by densitometry.

the impact of DUBA on NF-κB signaling appears to be confined to that triggered by RIG-I and MDA5.

In response to Sendai virus infection, which activates endogenous, cytosolic RIG-I (*2*), depletion of DUBA augmented secretion of IFN-A and RANTES in virus-infected HEK293 cells (Fig. 2B). DUBA depletion increased IFN-B secretion from the RAW264.3 macrophage cell line after treatment with ligands for TLR4 or TLR7 (Fig. 2C and fig. S3A) but did not affect TLR2-induced NF-κB–dependent production of IL-6. We used enforced expression of adaptors TRIF, IPS-1, and IL-1 receptor–associated kinase 4 (IRAK-4) to bypass activation of TLR3, RIG-I/MDA5, and TLRs 7 to 9, respectively, and to induce IFN-I expression. In each instance, depletion of DUBA enhanced activation of an IFN-A4 promoter (fig. S3B), supporting the notion that DUBA is a negative regulator of IFN-I expression downstream of multiple PRRs.

In HEK293 cells, ectopic expression of DUBA suppressed activation of IFN-A and IFN-B promoters but not a NF-κB–dependent promoter in cells treated with poly(I:C) to stimulate TLR3 (Fig. 2D). When coexpressed with RIG-I_{CARD} or $MDA5_{CARD}$, DUBA inhibited activation of a NF-κB–dependent promoter and the IFN-A and IFN-B promoters (Fig. 2D). All inhibitory effects of DUBA required its cysteine protease activity because the catalytic site mutant Cys^{224}→Ser^{224} (C224S) was inactive (*19*).

The downstream signaling machinery common to multiple PRRs that induce IFN-I expression includes the ubiquitin ligase TRAF3, the kinases TBK1 and IKKε, and the transcription factors IRF3 and IRF7 (*1*, *5*–*8*, *10*). DUBA reduction did not affect activation of IFN-B or IFN-A4 promoters in response to overexpressed IRF7 and IRF3, or to coexpressed TBK1 and IKKε (Fig. 2A and fig. S3C), indicating that DUBA must act upstream of TBK1 and IKKε. Consistent with these observations, overexpression of DUBA failed to inhibit IFN-A4 and IFN-B promoter activation by overexpressed IRF7 (Fig. 2D).

The conserved catalytic triad within the OTU domain of DUBA suggests that DUBA has deubiquitinating activity. Recombinant DUBA purified from *Escherichia coli* was incubated with Lys^{48}- or Lys^{63}-linked tetra-ubiquitin chains. DUBA degraded Lys^{63}-linked tetra-ubiquitin chains as effectively as did the positive control isopeptidase T (IsoT), but DUBA was not active against Lys^{48}-linked chains (Fig. 3A). The cysteine protease inhibitor *N*-ethylmaleimide (NEM) inhibited DUBA's ability to degrade Lys^{63}-linked ubiquitin chains, which is consistent with DUBA being a cysteine protease.

We immunoprecipitated FLAG epitope–tagged DUBA from HEK293 cells and identified associated proteins by mass spectrometry. Three peptides were from TRAF3 (Fig. 3B), which is a RING finger–type ubiquitin ligase that is essential for IFN-I expression downstream of multiple TLRs and the helicase receptors RIG-I and MDA5 (*6*, *7*). Increased binding of endogenous DUBA to endogenous TRAF3 was observed after poly(I:C) stimulation of TLR3 (Fig. 3C).

We hypothesized that DUBA might switch off IFN-I expression by deubiquitinating TRAF3. When overexpressed FLAG-TRAF3 was immunoprecipitated from lysates (after boiling in 1% SDS to remove associated proteins), smears corresponding to ubiquitinated TRAF3 were detected with hemagglutinin (HA)–tagged ubiquitin (Fig. 3D) and anti-TRAF3 (fig. S4). Ubiquitination of TRAF3 was also confirmed by the existence of ubiquitinated TRAF3 peptides in our mass spectrometry analysis. Wild-type DUBA, but not the catalytic site mutant DUBA C224S, reduced the amount of ubiquitinated TRAF3 detected. Sendai virus infection induced endogenous TRAF3 ubiquitination in fibroblasts, as evidenced by an anti-ubiquitin blot and slower migrating TRAF3 bands (fig. S5).

TRAF3 lacking the domain (ΔRING) required for E3 ligase activity and IFN-I production (*20*) exhibited less ubiquitin modification (Fig. 3D), which is consistent with an autoubiquitination mechanism. With the use of HA-tagged ubiquitin mutants in which only Lys^{48} or Lys^{63} was available to form polyubiquitin chains, TRAF3 mainly acquired Lys^{63}-linked polyubiquitin chains, and this modification was reduced by coexpression of DUBA (Fig. 3D). Ubiquitination of endogenous TRAF3 was detected upon overexpression of RIG-I_{CARD}, and this was increased by depletion of DUBA (Fig. 3E and fig. S6). Just as A20 is a DUB of TRAF6 in the classical NF-κB signaling pathway (*12*, *14*, *16*, *21*), DUBA appears to promote the removal of Lys^{63}-linked polyubiquitin chains from TRAF3. One possibility is that TRAF3 ubiquitination facilitates the recruitment of downstream signaling components. In support of this notion, DUBA overexpression partially reduced TRAF3-TBK1 interaction (Fig. 3F). Because TRAF3 is dispensable for RIG-I–mediated activation of NF-κB (*6*, *7*), the mechanism of how DUBA controls the NF-κB signaling remains to be elucidated.

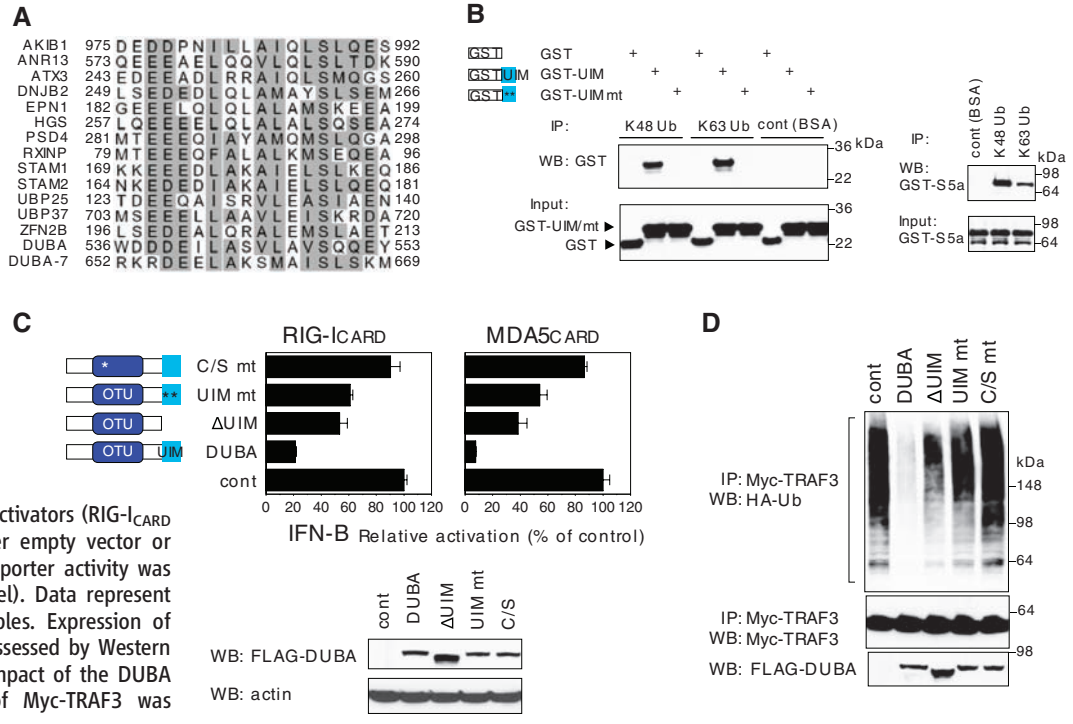


Fig. 4. Role for the UIM domain of DUBA. (**A**) Protein sequence alignment of UIM domain. Conserved amino acid residues are highlighted. (**B**) Interaction of the DUBA UIM domain with ubiquitin chains. GST, GST-UIM, and GST-UIM mt (L542A/S549A) proteins were incubated with agarose beads coated with either K48-Ub polyubiquitin-chains, K63-Ub polyubiquitin-chains, or bovine serum albumin (BSA). Bound material was analyzed by Western blotting with GST antibody (left panels). GST-S5a was used as a positive control (right panels). (**C**) Function of DUBA UIM. HEK293 cells were transfected with IFN-B reporter and indicated activators (RIG-I_{CARD} or $MDA5_{CARD}$), together with either empty vector or the version of DUBA indicated. Reporter activity was measured after 36 hours (top panel). Data represent the mean ± SD of triplicate samples. Expression of wild-type and mutant DUBA was assessed by Western blotting (bottom panel). (**D**) The impact of the DUBA mutants on the ubiquitination of Myc-TRAF3 was determined as in Fig. 3D.

Ubiquitin-binding domains (UBDs), such as the ubiquitin-associated domain (UBA) and ubiquitin-interacting motif (UIM), can influence various cellular events through binding to ubiquitinated proteins (*22*, *23*), but the function of UBDs in the deubiquitination reaction remains to be elucidated. DUBA has a putative UIM embedded in a conserved C-terminal helix (Fig. 4A). Both Lys^{48}- and Lys^{63}-linked polyubiquitin chains could bind glutathione *S*-transferase (GST) fused to the DUBA UIM in pull-down studies (Fig. 4B). Mutation of conserved residues in the DUBA UIM domain (L542A/S549A) prevented this binding. These results indicate that the DUBA UIM is capable of interacting with polyubiquitin chains. DUBA UIM mutants DUBAΔUIM and DUBA L542A/S549A were used to determine the role of the UIM in the negative regulation of IFN-I expression. Both mutants retained some ability to attenuate $RIG\text{-}I_{CARD}$–induced or $MDA5_{CARD}$-induced activation of an IFN-B promoter, although they were less effective than wild-type DUBA (Fig. 4C). As compared with wild-type DUBA, both UIM mutants were also consistently less effective at reducing the level of ubiquitinated TRAF3 in cells (Fig. 4D). These results suggest that there is an important role for the DUBA UIM in DUBA function but that the UIM may not be the sole substrate-recognition site. The binding affinities between ubiquitin and UIMs are typically very low (dissociation constant $K_d > 100$ μM) (*22*, *23*), so the DUBA UIM is predicted to play a supportive but dispensable role by capturing and presenting ubiquitinated substrates to the catalytic domain.

To determine whether DUBA, like A20, may function as part of a negative feedback loop, we asked whether DUBA was up-regulated in macrophages upon stimulation. When bone marrow–derived macrophages were exposed to lipopolysaccharide (LPS), there was substantial induction of DUBA protein (fig. S7). TRAF3 controls the nonclassical NF-κB/NF-κB2 pathway by regulating the NF-κB–inducing kinase–IKKα kinase cascade that results in the phosphorylation of NF-κB2/p100 and its subsequent proteasomal processing to transcriptionally competent p52 (*24*–*26*). Reduction of DUBA did not affect NF-κB2/p100 processing, either in the presence or absence of BAFF (B cell–activating factor belonging to the TNF family)–receptor engagement by B lymphocyte stimulator ligand (fig. S8). This finding suggests that DUBA affects only one aspect of TRAF3 signaling: namely, that which is required for IFN-I expression. Diseases such as systemic lupus erythematosus, where excess IFN-I production substantially contributes to pathology (*27*), stress the importance of negative regulation of IFN-I. Indeed, by suppressing IFN-I production, DUBA may function to inhibit the emergence of lupus-like autoimmune disorders.

References and Notes

1. S. Akira, S. Uematsu, O. Takeuchi, *Cell* **124**, 783 (2006).
2. H. Kato *et al.*, *Nature* **441**, 101 (2006).
3. A. G. Bowie, K. A. Fitzgerald, *Trends Immunol.* **28**, 147 (2007).
4. M. Yoneyama *et al.*, *Nat. Immunol.* **5**, 730 (2004).
5. E. Meylan, J. Tschopp, *Mol. Cell* **22**, 561 (2006).
6. H. Hacker *et al.*, *Nature* **439**, 204 (2006).
7. G. Oganesyan *et al.*, *Nature* **439**, 208 (2006).
8. K. Honda, A. Takaoka, T. Taniguchi, *Immunity* **25**, 349 (2006).
9. M. F. van den Broek, U. Muller, S. Huang, R. M. Zinkernagel, M. Aguet, *Immunol. Rev.* **148**, 5 (1995).
10. J. Hiscott, *J. Biol. Chem.* **282**, 15325 (2007).
11. C. M. Pickart, *Cell* **116**, 181 (2004).
12. Z. J. Chen, *Nat. Cell Biol.* **7**, 758 (2005).
13. S. M. Nijman *et al.*, *Cell* **123**, 773 (2005).
14. D. L. Boone *et al.*, *Nat. Immunol.* **5**, 1052 (2004).
15. G. R. Bignell *et al.*, *Nat. Genet.* **25**, 160 (2000).
16. A. Wullaert, K. Heyninck, S. Janssens, R. Beyaert, *Trends Immunol.* **27**, 533 (2006).
17. T. Kawai *et al.*, *Nat. Immunol.* **5**, 1061 (2004).
18. V. Hornung *et al.*, *Nat. Med.* **11**, 263 (2005).
19. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
20. S. K. Saha *et al.*, *EMBO J.* **25**, 3257 (2006).
21. B. Lamothe *et al.*, *J. Biol. Chem.* **282**, 4102 (2007).
22. L. Hicke, H. L. Schubert, C. P. Hill, *Nat. Rev. Mol. Cell Biol.* **6**, 610 (2005).
23. J. H. Hurley, S. Lee, G. Prag, *Biochem. J.* **399**, 361 (2006).
24. G. Bonizzi, M. Karin, *Trends Immunol.* **25**, 280 (2004).
25. G. Liao, M. Zhang, E. W. Harhaj, S. C. Sun, *J. Biol. Chem.* **279**, 26243 (2004).
26. J. Q. He *et al.*, *J. Exp. Med.* **203**, 2413 (2006).
27. V. Pascual, L. Farkas, J. Banchereau, *Curr. Opin. Immunol.* **18**, 676 (2006).

28. We thank members of the Dixit laboratory, I. Wertz, and Y.-M. Loo for discussion, K. Newton for proofreading, R. J. Deshaies for S5a, and M. Vasser, J.-A. Hongo, M. Hazen, P. Chang, A. P. Martinez, D. Davis, and J. Lee for technical assistance.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1145918/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References

31 May 2007; accepted 30 October 2007
Published online 8 November 2007;
10.1126/science.1145918
Include this information when citing this paper.

Sensing X Chromosome Pairs Before X Inactivation via a Novel X-Pairing Region of the *Xic*

S. Augui,[1] G. J. Filion,[1] S. Huart,[1] E. Nora,[1] M. Guggiari,[1] M. Maresca,[2] A. F. Stewart,[2] E. Heard[1]*

Mammalian dosage compensation involves silencing of one of the two X chromosomes in females and is controlled by the X-inactivation center (*Xic*). The *Xic*, which includes *Xist* and its antisense transcription unit *Tsix/Xite*, somehow senses the number of X chromosomes and triggers *Xist* up-regulation from one of the two X chromosomes in females. We found that a segment of the mouse *Xic* lying several hundred kilobases upstream of *Xist* brings the two *Xic*s together before the onset of X inactivation. This region can autonomously drive *Xic* trans-interactions even as an ectopic single-copy transgene. Its introduction into male embryonic stem cells is strongly selected against, consistent with a possible role in trans-activating *Xist*. We propose that homologous associations driven by this novel X-pairing region (*Xpr*) of the *Xic* enable a cell to sense that more than one X chromosome is present and coordinate reciprocal *Xist/Tsix* expression.

In female mammals, random X-chromosome inactivation (XCI) is triggered during early development. The *Xic*, which contains the *Xist* gene and its antisense unit *Tsix/Xite*, controls the initial steps of XCI. These steps include sensing the number of X chromosomes, counting the X:autosome ratio, and choosing the X to inactivate. The *Xic* was originally defined as a region of 680 kb to 1.2 Mb that was essential for XCI [reviewed in (*1*)], but so far no single-copy *Xist*-containing transgene has been found to recapitulate the random XCI process in a manner equivalent to that found in X-autosome translocations (*2*). Complete *Xic* function therefore requires long-range elements in addition to *Xist/Tsix/Xite*.

At the onset of XCI, *Xist* becomes up-regulated and *Tsix* down-regulated on the future inactive X (*3*). Although antisense transcription negatively regulates *Xist* in cis (*4*–*6*), there must also be some form of trans-regulation between the two *Xist/Tsix* loci, given the tight coordination of their mutually exclusive expression patterns. Recent insight into this comes from the finding that the two *Xic*s associate transiently at around the time that *Xist* is monoallelically up-regulated (*7*, *8*). Pairing is disrupted when *Tsix* or *Xite* are deleted (*7*, *8*). CTCF and transcription are thought to participate in pairing (*9*). However, although ectopic multicopy arrays of *Tsix* and *Xite* can associate in trans with the endogenous *Xic* (*8*), single-copy *Xist/Tsix/Xite* transgenes such

[1]CNRS UMR218, Curie Institute, 26 rue d'Ulm, Paris 75005, France. [2]BioInnovationsZentrum, Technical University Dresden, Tatzberg 47-51, D-01307 Dresden, Germany.

*To whom correspondence should be addressed. E-mail: edith.heard@curie.fr

as YAC PA2 (Fig. 1A) cannot (*7*); hence, additional *Xic* sequences are normally required for trans-interaction.

To identify the missing functional elements of the *Xic*, we performed a systematic search for homologous trans-interacting regions by DNA fluorescence in situ hybridization (FISH) in differentiating PGK12.1 female embryonic stem (ES) cells, with the use of bacterial artificial chromosome (BAC) probes spanning the 800-kb murine *Xic* candidate region (Fig. 1A) (*10*). Interallelic *Xic* distances were measured in undifferentiated (day 0) and differentiating ES cells, during the critical time window in which *Xist* becomes monoallelically up-regulated (Fig. 1B).

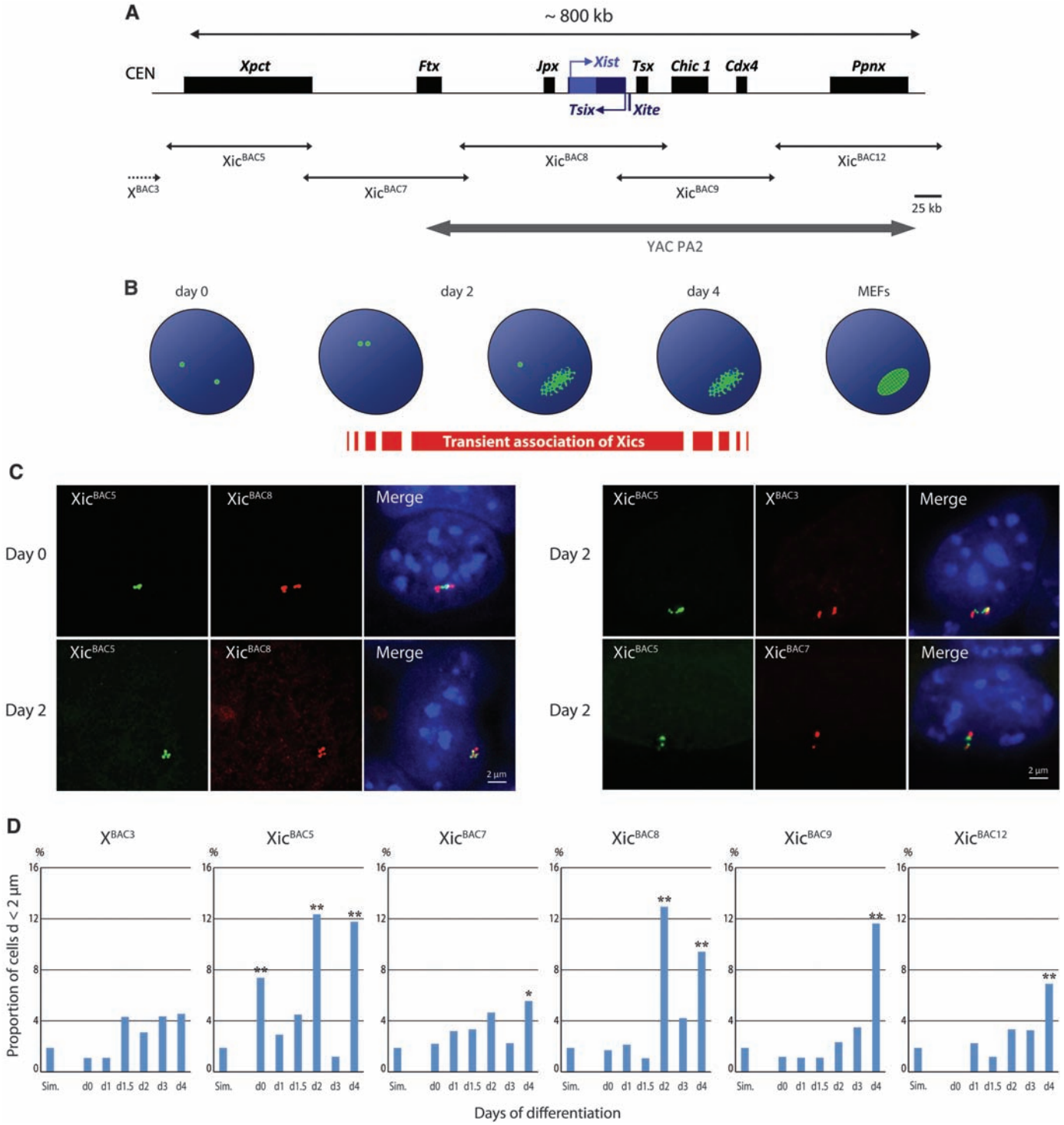


Fig. 1. Evidence for a new *Xic* region involved in homologous trans-interactions. (**A**) Map of the *Xic* region. (**B**) Timing of events during female ES cell differentiation. (**C**) Two-color DNA FISH using *Xic* probes (green or red, as indicated) on undifferentiated and differentiating PGK12.1 XX wild-type cells; DAPI (4′,6′-diamidino-2-phenylindole) is shown in blue. Maximum-intensity projections of 3D image stacks (two or three planes) are shown. (**D**) Binned interallelic distances of ≤2 μm (3D measurements) are shown for each probe at each stage of differentiation. Statistically significant pairing frequencies (*11*) are indicated: *$P < 0.05$, **$P < 0.001$ ($n > 80$ per day of differentiation). Four independent differentiations showed similar kinetics. Note that Xic^{BAC5} proximity in undifferentiated ES cells, although significant, varied between experiments and may in fact represent the very earliest differentiation stages.

As expected (*7*, *8*), the *Xist*/*Tsix* region (Xic^{BAC8}) showed statistically significant peaks of interallelic proximity at day 2 of differentiation ($P < 0.001$) (*11*), which corresponds to the onset of monoallelic *Xist* up-regulation, and at day 4 as well (Fig. 1, C and D, and fig. S1), but not in mouse embryonic fibroblasts, which represent a post-XCI stage (fig. S1). Of the other BACs tested, only the most proximal part of the *Xic* (Xic^{BAC5}) showed significantly short ($P < 0.001$) interallelic distances at days 2 and 4 of differentiation, similarly to Xic^{BAC8}. Indeed, Xic^{BAC5} proximity was observed even in undifferentiated and early-differentiating ES cells (Fig. 1, C and D, and fig. S1). Such early trans-association was unique to the Xic^{BAC5} region, as within the same time course, significant Xic^{BAC8} interallelic proximity was found only at days 2 and 4 (Fig. 1, C and D, and fig. S1).

Our data define a proximal region of the *Xic* that shows interallelic associations before the onset of XCI. The 136-kb Xic^{BAC5} segment lies 200 kb 5′ to *Xist*, within a heterochromatic "hotspot" (*12*, *13*) and overlapping with the *Xpct* (*Slc16a2*) gene (*11*). The early onset of interallelic Xic^{BAC5} associations indicated that they occurred independently of Xic^{BAC8} pairing. To test this, we analyzed D102 ($XX^{\Delta 65kb}$) ES cells, which are deleted for *Tsix*/*Xite* on one X, and in which Xic^{BAC8} pairing is abolished at day 2 (*7*). Counting and choice are affected by this deletion, but not sensing, as XCI still occurs (*14*). Statistically significant proximity between the two Xic^{BAC5} regions (fig. S2) was found throughout differentiation in the D102 line. Interallelic Xic^{BAC5} associations were similarly not disrupted in XX ES cells carrying a *Xist* deletion on one allele, in which sensing, counting, and choice are still functional (*15*) (fig. S3). Thus, Xic^{BAC5} trans-associations occur independently of the *Xist*/*Tsix*/*Xite* (Xic^{BAC8}) region during ES cell differentiation, and Xic^{BAC5} may be critical for bringing the two *Xic*s together.

We investigated whether these apparently transient *Xic* interactions might be dependent on the cell cycle. DNA FISH on ES cells sorted by fluorescence-activated cell sorting (FACS) or after proliferating cell nuclear antigen (PCNA) staining (Fig. 2, A and B) revealed that *Xic* associations occur during S phase, in cells with an early S-phase pattern (*16*). When in close proximity, the *Xic* loci did not overlap with PCNA foci, and two chromatids per allele were always seen. Thus, *Xic* associations occur at a precise cell cycle stage, after replication and sister chromatid separation (Fig. 2C and fig. S4) (*11*).

We also investigated the manner in which the *Xic*s come together in the nucleus. Whenever the two Xic^{BAC5} loci were juxtaposed, the two X-chromosome territories were found to be intimately associated (95%, $n = 20$) (Fig. 2D). However, the frequencies with which the two Xs were associated were not significantly different from those for autosomes in ES cells (fig. S5B). Thus, random collision of the two X chromosomes may be necessary for interallelic *Xic* association. However, it cannot be sufficient, as regions adjacent to Xic^{BAC5} (Fig. 1D) did not show significant interallelic proximity during early differentiation, implying some intrinsic property of Xic^{BAC5} in establishing *Xic* trans-interactions.

To address this possibility directly, we created ES cell lines carrying Xic^{BAC5} transgenes. Parallel transfections using either Xic^{BAC5} or Xic^{BAC8} were performed in several different ES cell lines (Fig. 3A). Transgenic female (PGK12.1) ES cells were obtained with both BACs, although Xic^{BAC5} clone frequency was always lower than for Xic^{BAC8} (Fig. 3A). Single-copy transgenes appeared to be present in three Xic^{BAC5} transgenic clones analyzed (fig. S6). In all three clones, the transgene showed significant trans-association ($P < 0.001$) with the endogenous *Xic* (Fig. 3, B and C). Endogenous *Xic*-*Xic* interactions were reduced but still detectable. Overall frequencies of *Xic* interallelic associations in transgenic clones were in fact higher than in wild-type XX cells (Fig. 3C). Trans-interactions occurred in a pairwise fashion (Tg-*Xic* or *Xic*-*Xic*) (Fig. 3B), and the three loci (Tg-*Xic*-*Xic*) were rarely (<1%, $n > 300$) found together simultaneously. Transgenic lines made with a control BAC (X^{BAC3}) showed no significant interallelic *Xic*-transgene proximity (<3%, $n > 80$). DNA FISH with chromosome paint probes revealed that the transgenic and endogenous loci associate when the autosome and the X are in close proximity (63%, $n = 20$) (Fig. 3D) and that a shift of the loci to the outer edge of their territories occurs during this period (fig. S5D). In summary, Xic^{BAC5} appears to have an inherent capacity to undergo homologous trans-interactions autonomously. We thus term this locus the "X-pairing region" (*Xpr*).

We also assessed whether the presence of an *Xpr* transgene in XX ES cells affected the onset of XCI. Two clones (9 and 25) displayed early *Xist* RNA accumulations in up to 30% of undifferentiated cells, a situation not found in the parental PGK12.1 line (fig. S7). These clones also showed a slightly higher frequency of *Xist* RNA accumulation during differentiation than in PGK12.1 cells. Thus, an ectopic copy of *Xpr* may actually enhance the capacity of XX ES cells to induce *Xist* RNA, at least in two of the three clones examined. We were

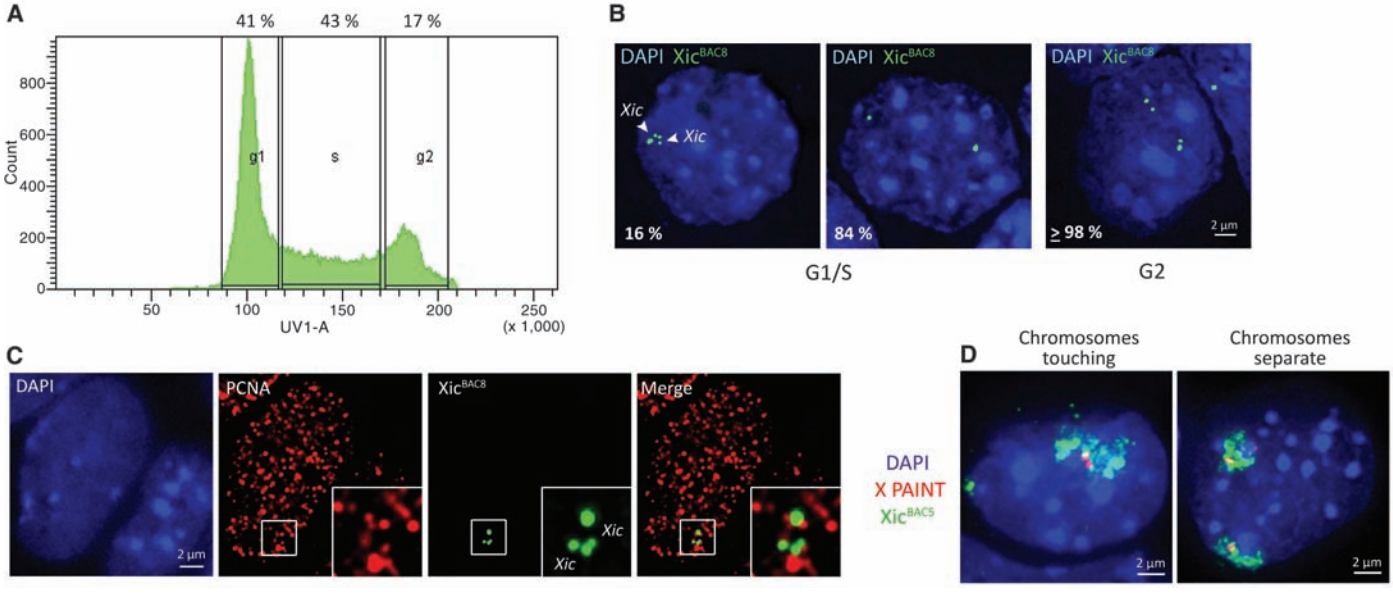
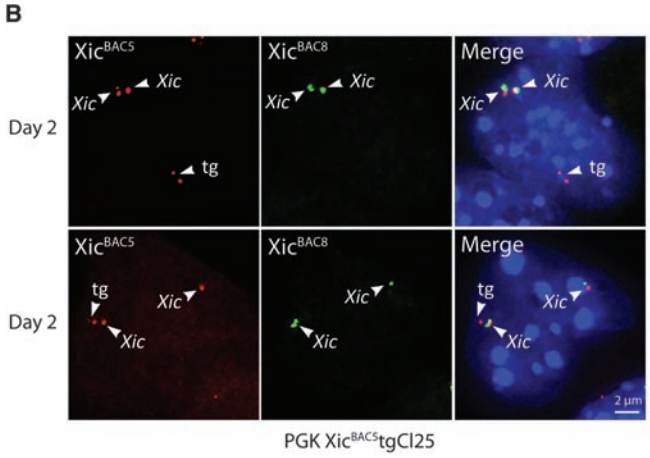
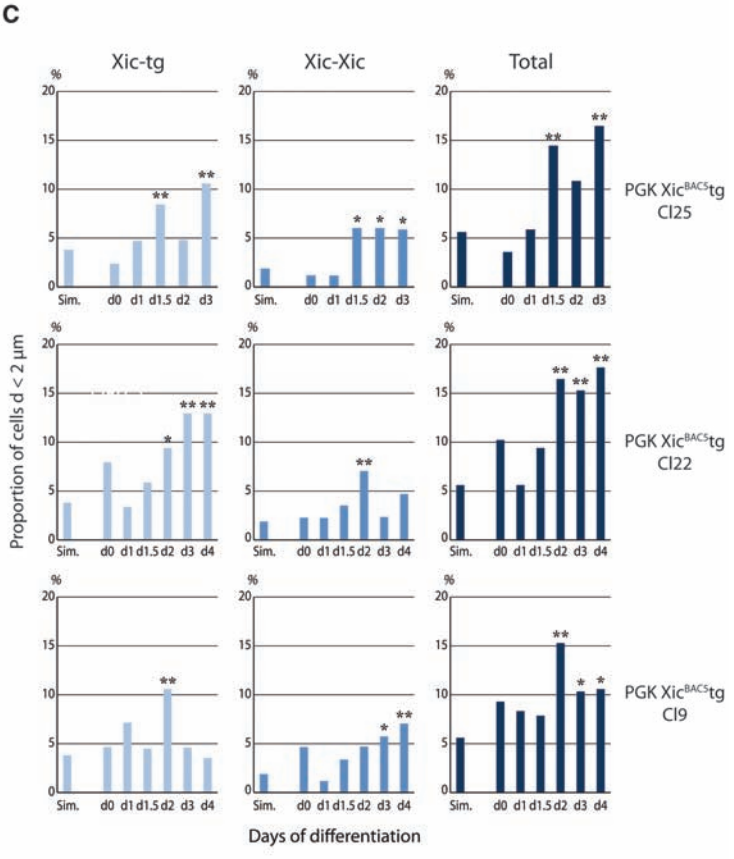


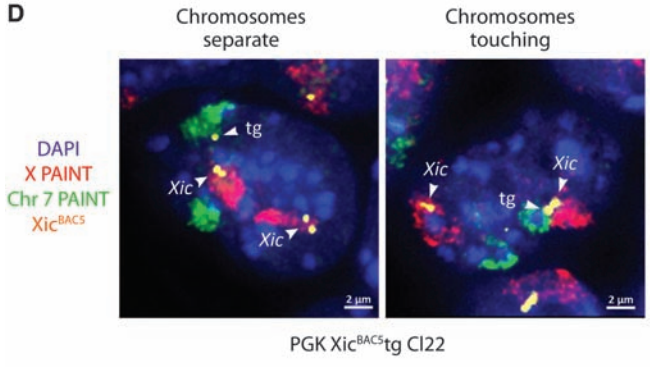
Fig. 2. Nuclear and cell cycle dynamics of *Xic* associations. (**A**) FACS-sorted PGK12.1 (XX wild-type) cells differentiated for 2 days. The proportion of cells obtained at each stage of the cell cycle (G_1, S, and G_2/M) is indicated above the graph. (**B**) DNA FISH (Xic^{BAC8}) on sorted cells shows that *Xic* proximity occurs in the G_1/S population. Two chromatids per locus are detectable during trans-association (arrowheads) ($n = 120$). (**C**) PCNA immunofluorescence (red) combined with Xic^{BAC8} DNA FISH (green) shows that *Xic* associations occur in cells with an early S-phase pattern. (**D**) DNA FISH (X paint, green; Xic^{BAC5} probe, red; DAPI, blue) shows that Xic^{BAC5} pairing involves intimate association between the two X chromosomes.

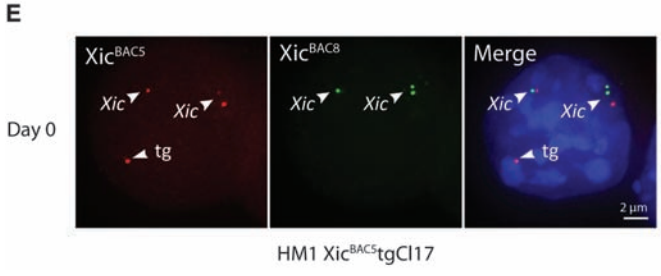
A

ES cell line	PGK (XX)	HM1 (XY)	CK35 (XY)	CK35 (XY)	CK35 (XY)	CK35 (XY)	PGK (XX)	PGK (XX)	E14 (XY)	E14 (XY)
Transfected DNA	Xic^{BAC5} + $Hygro^R$	Xic^{BAC5} + $Hygro^R$	Xic^{BAC5} + $Hygro^R$	$Hygro^R$	$Xic^{BAC5}Bsd^R$	$Xic^{BAC8}Bsd^R$	$Xic^{BAC5}Bsd^R$	$Xic^{BAC8}Bsd^R$	$Xic^{BAC5}Bsd^R$	$Xic^{BAC8}Bsd^R$
No. resistant clones isolated	45	18	0	>50	3	26	3	20	9	37
No. clones with Xic^{BAC5}tg	5	1 (XXY)	0	NA	0	NA	NT	NA	1	NA









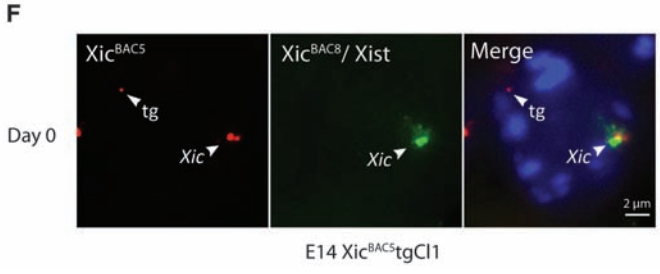


Fig. 3. Xic^{BAC5} transgenesis assays. (**A**) Summary of clone numbers obtained by stable transfection with Xic^{BAC5} and Xic^{BAC8} with different ES cell lines and selectable markers. Parallel experiments are grouped in columns within double lines. NA, not applicable (clones not analyzed); NT, not tested. (**B**) Representative DNA FISH analysis of a PGK12.1 female Xic^{BAC5} transgenic line, clone 25, shows trans-association between endogenous *Xic*s (upper panel) or between an endogenous *Xic* and the Xic^{BAC5} transgene (lower panel). Images are maximum-intensity projections of 3D image stacks (two or three planes). (**C**) Binned distances of ≤2 μm (3D measurements) reveal the kinetics of Xic^{BAC5} associations during ES cell differentiation in clones 9, 22, and 25. Numbers of trans-associations (*Xic-Xic*, *Xic*-Tg, or total) are shown; statistically significant pairing frequencies are indicated: $^*P < 0.05$, $^{**}P < 0.001$ ($n > 80$ per day of differentiation). (**D**) Triple-probe DNA FISH (X paint, green; chromosome 7 paint, red; Xic^{BAC5} probe, yellow) shows endogenous-transgenic Xic^{BAC5} pairing with intimate association between the X chromosome and the transgenic autosome in clone 22. (**E**) DNA FISH (Xic^{BAC5}/Xic^{BAC8}) reveals two endogenous *Xic*s in clone 17, the HM1 Xic^{BAC5} transgenic clone. (**F**) RNA/DNA FISH shows early *Xist* RNA accumulations in undifferentiated cells of clone 1, the E14 Xic^{BAC5} transgenic clone.

unable to analyze the behavior of *Xpr* transgenes in an XY context because hardly any transgenic clones could be generated with Xic^{BAC5}, despite numerous attempts (Fig. 3A). This suggests a strong counterselection against the presence of this region in an XY context, which could be due to an effect of the transgene and/or to epigenetic differences between XX and XY ES cells (*17*). The single HM1 transgenic clone obtained appeared to be XXY (Fig. 3E). The single E14 transgenic clone obtained rapidly lost its transgene upon passaging and differentiation. In early passages, this clone displayed *Xist* RNA accumulation on the single X in a small proportion of undifferentiated ES cells (Fig. 3F), a situation not observed in wild-type male ES cells. *Xist* RNA accumulation on the X in XY cells would rapidly lead to counterselection owing to X-chromosome nullisomy. In an XX (or XXY) context, counterselection might be partly avoided as a result of the second X chromosome. The apparent up-regulation of *Xist* by Xic^{BAC5} transgenes in undifferentiated ES cells was unexpected, as *Xist* is normally repressed in such cells through unknown mechanisms (*3*). Our results suggest that ectopic copies of *Xpr* somehow alleviate *Xist* repression in undifferentiated ES cells, although clearly this is normally tightly controlled when *Xpr* is at its endogenous location. We propose that when two or more *Xpr* copies are present in a cell, this relieves *Xist* repression or activates *Xist* in trans in a transvection-like process (*18*, *19*). This may represent part of the sensing step that triggers XCI in cells with more than one X chromosome (Fig. 4A). A prediction of this is that in XX ES cells, *Xpr* trans-associations could symmetrically activate *Xist*. Indeed, close examination of early-differentiating XX ES cells revealed a significant percentage (15%) of cells with biallelic *Xist* RNA clusters, peaking at day 1.5 (Fig. 4B), in a time window between *Xpr* and *Tsix* trans-associations.

Our data reveal a region of the *Xic* that participates in pairwise trans-sensing of X chromosomes. *Xpr* pairing may represent the previously proposed X-linked competence factor, present only in XX cells and required for initiation of XCI (*20*). We propose that *Xpr* pairing triggers *Xist* up-regulation from both X chromosomes but also rapidly renders this biallelic expression monoallelic by enabling *Tsix/Xite* pairing (Fig. 4A). Future studies will define the elements mediating these *Xpr* functions and will determine whether the heterochromatic nature of this region underlies its capacity to associate in trans (*18*, *19*).

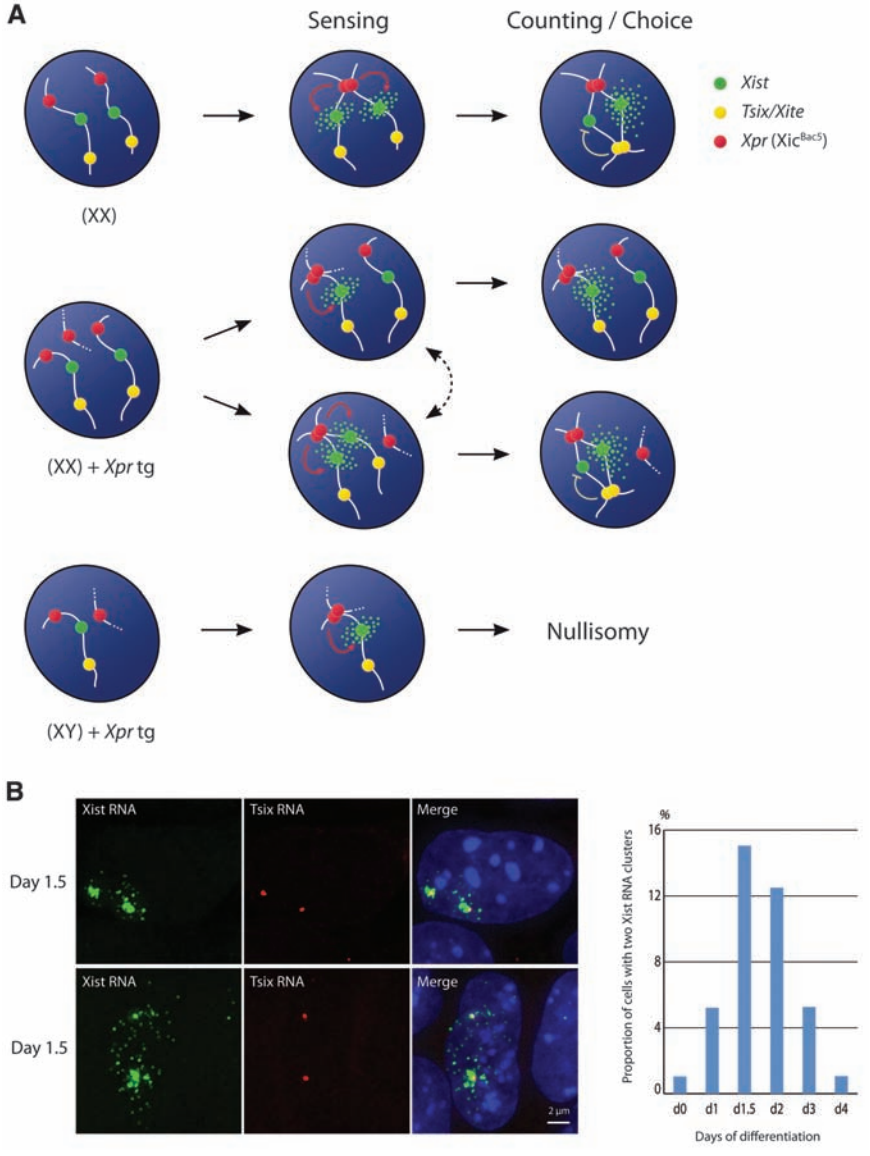


Fig. 4. Early steps in the initiation of X inactivation. (**A**) Model. *Xic* pairing is induced by the newly defined X-pairing region (*Xpr*, red) and could represent the trigger for XCI by up-regulating *Xist* (green) as part of a sensing step. *Xpr* pairing may also facilitate interallelic association between *Tsix/Xite* regions (yellow), which results in reciprocal *Xist/Tsix* regulation, rendering *Xist* up-regulation monoallelic. Ectopic copies of the *Xpr* may trigger *Xist* up-regulation; in the absence of a linked *Tsix* region, this leads to lethality through functional X-chromosome nullisomy. In females, *Xpr* transgenes may be better tolerated because of the presence of the second X chromosome, but *Xist* up-regulation seems to be enhanced. (**B**) Left panel: *Xist/Tsix* RNA FISH (*Xist* probe, green; *DXPas34* probe, red) reveals biallelic *Xist* RNA clusters and *Tsix* expression during early differentiation. Right panel: Biallelic *Xist* expression peaks just before *Tsix/Xite* association, as assessed by Xic^{BAC8} DNA FISH within the same ES cell differentiation (n = 100 per day of differentiation).

References and Notes

1. E. Heard, P. Avner, *Hum. Mol. Genet.* **3**, 1481 (1994).
2. E. Heard, F. Mongelard, D. Arnaud, P. Avner, *Mol. Cell. Biol.* **19**, 3156 (1999).
3. J. T. Lee, N. Lu, *Cell* **99**, 47 (1999).
4. S. Luikenhuis, A. Wutz, R. Jaenisch, *Mol. Cell. Biol.* **21**, 8512 (2001).
5. N. Stavropoulos, N. Lu, J. T. Lee, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10232 (2001).
6. T. B. Nesterova *et al.*, *Genes Dev.* **17**, 2177 (2003).
7. C. P. Bacher *et al.*, *Nat. Cell Biol.* **8**, 293 (2006).
8. N. Xu, C.-L. Tsai, J. T. Lee, *Science* **311**, 1149 (2006); published online 18 January 2006 (10.1126/science.1122984).
9. N. Xu, M. E. Donohoe, S. S. Silva, J. T. Lee, *Nat. Genet.* **39**, 1390 (2007).
10. C. Chureau *et al.*, *Genome Res.* **12**, 894 (2002).
11. See supporting material on *Science* Online.
12. E. Heard *et al.*, *Cell* **107**, 727 (2001).
13. C. Rougeulle *et al.*, *Mol. Cell. Biol.* **24**, 5475 (2004).
14. P. Clerc, P. Avner, *Nat. Genet.* **19**, 249 (1998).
15. G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, *Nature* **379**, 131 (1996).
16. H. Leonhardt *et al.*, *J. Cell Biol.* **149**, 271 (2000).
17. I. Zvetkova *et al.*, *Nat. Genet.* **37**, 1274 (2005).
18. S. Henikoff, L. Comai, *Cell* **93**, 329 (1998).
19. V. Pirrotta, *Biochim. Biophys. Acta* **1424**, M1 (1999).
20. J. T. Lee, *Science* **309**, 768 (2005).

21. We thank P. Le Baccon for microscopy, C. Picard for technical assistance, members of UMR218 for stimulating discussions, P. Avner and P. Clerc for D102 cells, and N. Brockdorff for XT67E1 cells. Supported by the Fondation pour la Recherche Médicale, the Human Frontier Science Program Organization, the European Network of Excellence (Epigenome), the High-Throughput Epigenetic Regulatory Organization in Chromatin (HEROIC) Integrated Project (grant LSHG-CT-2005-018883), and the Schlumberger Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1632/DC1
Materials and Methods
Figs. S1 to S7
References

20 August 2007; accepted 5 November 2007
10.1126/science.1149420

Orchestration of the DNA-Damage Response by the RNF8 Ubiquitin Ligase

Nadine K. Kolas,[1]* J. Ross Chapman,[2]* Shinichiro Nakada,[1]* Jarkko Ylanko,[1,3] Richard Chahwan,[2] Frédéric D. Sweeney,[1,3] Stephanie Panier,[1] Megan Mendez,[1] Jan Wildenhain,[1] Timothy M. Thomson,[4] Laurence Pelletier,[1,3] Stephen P. Jackson,[2]† Daniel Durocher[1,3]†

Cells respond to DNA double-strand breaks by recruiting factors such as the DNA-damage mediator protein MDC1, the p53-binding protein 1 (53BP1), and the breast cancer susceptibility protein BRCA1 to sites of damaged DNA. Here, we reveal that the ubiquitin ligase RNF8 mediates ubiquitin conjugation and 53BP1 and BRCA1 focal accumulation at sites of DNA lesions. Moreover, we establish that MDC1 recruits RNF8 through phosphodependent interactions between the RNF8 forkhead-associated domain and motifs in MDC1 that are phosphorylated by the DNA-damage activated protein kinase ataxia telangiectasia mutated (ATM). We also show that depletion of the E2 enzyme UBC13 impairs 53BP1 recruitment to sites of damage, which suggests that it cooperates with RNF8. Finally, we reveal that RNF8 promotes the G_2/M DNA damage checkpoint and resistance to ionizing radiation. These results demonstrate how the DNA-damage response is orchestrated by ATM-dependent phosphorylation of MDC1 and RNF8-mediated ubiquitination.

DNA double-strand breaks (DSBs) are highly cytotoxic lesions, and to ensure that they are repaired with minimal impact on genome stability, cells mount a complex DNA-damage response (DDR) that includes the spatial reorganization of DSB repair and signaling proteins into subnuclear structures—ionizing radiation–induced foci (IRIF)—that surround DSB sites (*1*, *2*). Most IRIF formation depends on phosphorylation of the histone variant H2AX (to form γH2AX) by the DNA-dependent and ataxia telangiectasia mutated (ATM) protein kinases (*3*–*6*). The γH2AX epitope is bound by MDC1 (*7*–*10*) that then promotes IRIF formation by other proteins, including 53BP1, Nijmegen-breakage-syndrome protein NBS1, and BRCA1 (*11*, *12*). BRCA1 recruitment to IRIF requires

[1]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Ontario, Canada. [2]The Wellcome Trust and Cancer Research UK Gurdon Institute, and the Department of Zoology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. [3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [4]Department of Molecular and Cellular Biology, Instituto de Biología Molecular de Barcelona calle Jordi Girona 18-26, 08034 Barcelona, Spain.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: durocher@mshri.on.ca (D.D.); s.jackson@gurdon.cam.ac.uk (S.P.J.)

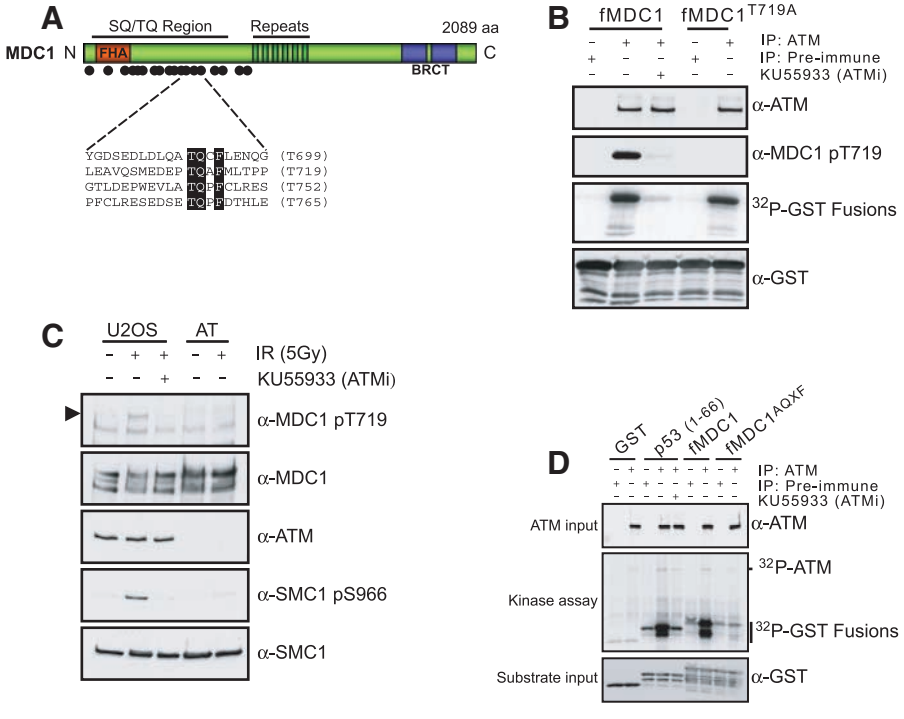


Fig. 1. The MDC1 TQXF motifs are ATM targets required for 53BP1 IRIF. (**A**) Domain architecture of MDC1, with ATM consensus sites (dots). (**B**) MDC1 T719 is phosphorylated by ATM in vitro. GST-MDC1$^{679\text{-}778}$ (fMDC1) or GST-MDC1$^{679\text{-}778\text{-}T719A}$ (fMDC1^{T719A}) were incubated with antibody to ATM or preimmune complexes in the presence or absence of the ATM kinase inhibitor KU55933. (**C**) MDC1 TQXF motifs are phosphorylated by ATM in vivo. Lysates from U2OS or AT22IJE (AT) cells were immunoblotted with the indicated antibodies. The arrowhead points to phospho-MDC1. (**D**) The MDC1 TQXF motifs are phosphorylated by ATM in vitro. Kinase reactions with antibody to ATM or preimmune complexes and the following substrates: GST, GST-p53$^{1\text{-}66}$ [p53(1-66)], fMDC1 or GST-MDC1$^{679\text{-}778\text{-}AQXF}$ (fMDC1^{AQXF}), as in (B). (**E** and **F**) The TQXF cluster is required for 53BP1 IRIF. U2OS cells expressing siRNA-resistant GFP-MDC1 (MDC1*) or GFP-MDC1^{AQXF} (MDC1^{*AQXF}) were transfected with siRNA against MDC1 (siMDC1) or luciferase (siCTRL), and after irradiation (5 Gy) were stained for MDC1 and 53BP1 (E) and quantitated (F) ($N = 4 \pm$ SD).

its interaction with the ubiquitin-binding protein RAP80 (*13*–*16*) that interacts with lysine 63 (K63)–linked polyubiquitinated protein(s) at sites of DNA damage (*15*). Here, we identify RNF8 as the prime ubiquitin ligase for ubiquitination at DSB sites, define its functional importance in the DDR, and establish how RNF8 is recruited to sites of DNA damage through interactions with MDC1.

MDC1 is phosphorylated in an ATM-dependent manner in response to ionizing radiation (IR) (*11*, *12*). Potential ATM target sites (consensus S/T-Q) cluster in the MDC1 N terminus, the most notable being four adjacent motifs conforming to the consensus TQXF (Fig. 1A and fig. S1). Notably, antibodies raised against a peptide encoding phospho-T719 (fig. S2A) indicated that it is targeted by ATM in vitro (Fig. 1B) and in vivo (Fig. 1C and fig. S2B). However, in vitro assays with bacterially expressed MDC1 fragments revealed that T719 was not the only site of ATM modification. Phosphorylation was only abolished when an "AQXF" mutant protein bearing threonine-to-alanine substitutions in all four TQXF motifs was used as substrate (Fig. 1D). These data and the recent identification of another TQXF site (T752) as an ATM target (*17*) therefore imply that MDC1 TQXF motifs are likely all modified by ATM and may function redundantly with one another.

To address the function of the MDC1 TQXF motifs, we used small interfering RNA (siRNA) to deplete endogenous MDC1 in human U2OS cells stably expressing siRNA-resistant wild-type MDC1 or the MDC1 AQXF mutant. Although both wild-type and AQXF mutant proteins formed IRIF and supported IRIF formation by NBS1 (fig. S3), only wild-type MDC1 promoted effective IRIF formation by 53BP1 (Fig. 1, E and F), revealing that ATM-mediated phosphorylation of MDC1 facilitates 53BP1 focus formation. Indeed, although 53BP1 IRIF formation can occur in ATM-deficient AT cells (*18*, *19*), we found that pharmacological inhibition of ATM impaired 53BP1 focus formation (fig. S4). Although these results suggested that 53BP1 might directly bind the phosphorylated TQXF motifs, we were unable to detect such interactions, which suggests that MDC1-dependent 53BP1 IRIF formation is likely mediated by an additional factor.

To uncover proteins that regulate 53BP1 focus formation, we mined an ongoing RNA interference (RNAi) screen that employs 53BP1 focus formation as a readout (Fig. 2A and fig. S5, A to C). Notably, the three siRNAs that most potently impaired 53BP1 focus formation targeted transcripts encoding MDC1, ubiquitin, and the RNF8 E3 ubiquitin ligase that was recently shown to control mitosis (*20*) (Fig. 2A). Notably, we found that depletion of RNF8 by siRNA or enzyme-generated siRNA pools (esiRNAs) (*21*) abrogated 53BP1 foci while preserving MDC1 IRIF, thus phenocopying the MDC1^{AQXF} mutation (Fig. 2, B and C, and fig. S6A). Introduction of RNAi-resistant murine RNF8 into cells transfected with RNF8 siRNA restored 53BP1 focus formation (Fig. 2D and fig. S7), indicating the specificity of RNF8 depletion and confirming that RNF8 promotes 53BP1 IRIF formation.

RNF8 possesses an N-terminal forkhead-associated (FHA) domain (*22*) and a C-terminal RING-finger domain responsible for its ubiquitin ligase activity (Fig. 2E) (*23*). By complementing the human RNF8 RNAi phenotype with siRNA-resistant murine RNF8, we established that mutations in either the FHA domain (RNF8^{R42A}) or RING-finger domain (RNF8^{C406S}) abrogated the ability of RNF8 to support 53BP1 IRIF formation (Fig. 2, E and F, and fig. S7). As FHA domains bind phosphothreonine-bearing epitopes, the above data suggested that RNF8 might interact with ATM-phosphorylated MDC1. Indeed, the RNF8 FHA domain, but not an FHA-domain mutant (RNF8^{R42A}), bound specifically and in a phospho-dependent manner to TQXF peptides corresponding to MDC1 Thr719 or Thr752 (Fig. 3A). Furthermore, these phosphorylated TQXF peptides retrieved RNF8 from HeLa nuclear extracts, whereas the corresponding unphosphorylated peptides did not (Fig. 3B). Notably, epitope-tagged RNF8, but not RNF8^{R42A}, was detected in MDC1 immunoprecipitates in a manner that was enhanced by irradiation (Fig. 3C), which suggests that MDC1 might recruit RNF8 to sites of DNA damage. To test this idea, we generated cell lines stably expressing RNF8 fused to yellow fluorescent protein (YFP). As shown by analysis of live cells (Fig. 3D) and fixed samples (fig. S8), addition of the radio-mimetic drug phleomycin or

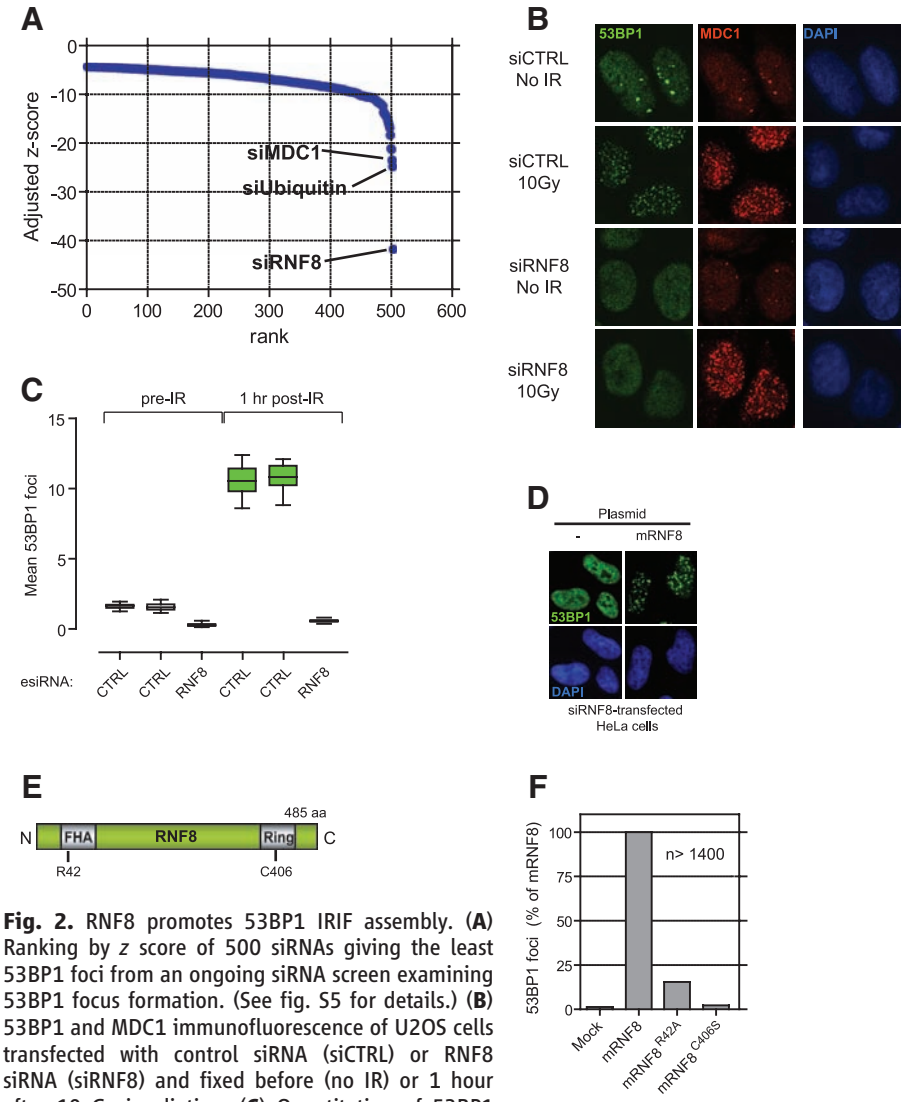


Fig. 2. RNF8 promotes 53BP1 IRIF assembly. (**A**) Ranking by *z* score of 500 siRNAs giving the least 53BP1 foci from an ongoing siRNA screen examining 53BP1 focus formation. (See fig. S5 for details.) (**B**) 53BP1 and MDC1 immunofluorescence of U2OS cells transfected with control siRNA (siCTRL) or RNF8 siRNA (siRNF8) and fixed before (no IR) or 1 hour after 10 Gy irradiation. (**C**) Quantitation of 53BP1 IRIF in HeLa cells transfected with the indicated esiRNAs (CTRL against luciferase). *N* = 16, data displayed using box-and-whisker plots. (**D**) Transfection of siRNA-resistant murine *RNF8* in HeLa cells restores 53BP1 IRIF formation caused by RNF8 depletion. (**E**) Domain architecture of RNF8. Numbering refers to murine RNF8. (**F**) Rescue of RNF8 depletion by murine RNF8 but not the FHA-(R42A) or RING finger–mutated (C406s) mutants. 53BP1 foci were quantitated 1 hour after 10 Gy irradiation in siRNF8-treated cells. Data for wild-type RNF8 were set at 100%. More than 1400 cells per condition were counted.

irradiation resulted in the accumulation of YFP-RNF8 into foci that colocalized with γH2AX. In support of such events being brought about by ATM-mediated phosphorylation of MDC1, YFP-RNF8 recruitment was impaired by the selective ATM inhibitor KU55933 (*24*) (Fig. 3D).

As was the case for MDC1 foci, we found that γH2AX, NBS1, and FANCD2 foci formed efficiently in RNF8-depleted cells (fig. S9). In contrast, BRCA1 IRIF formation was impaired upon RNF8 depletion (Fig. 3E). BRCA1-interacting protein RAP80 is required for BRCA1 IRIF formation (*13*–*16*), and RAP80 itself forms IRIF in a manner that involves interactions between its ubiquitin-interacting motif (UIM) and K63-linked ubiquitin chains at sites of DNA lesions (*13*–*16*). Because RNF8 is a ubiquitin ligase that promotes BRCA1 IRIF, we speculated that RNF8 might

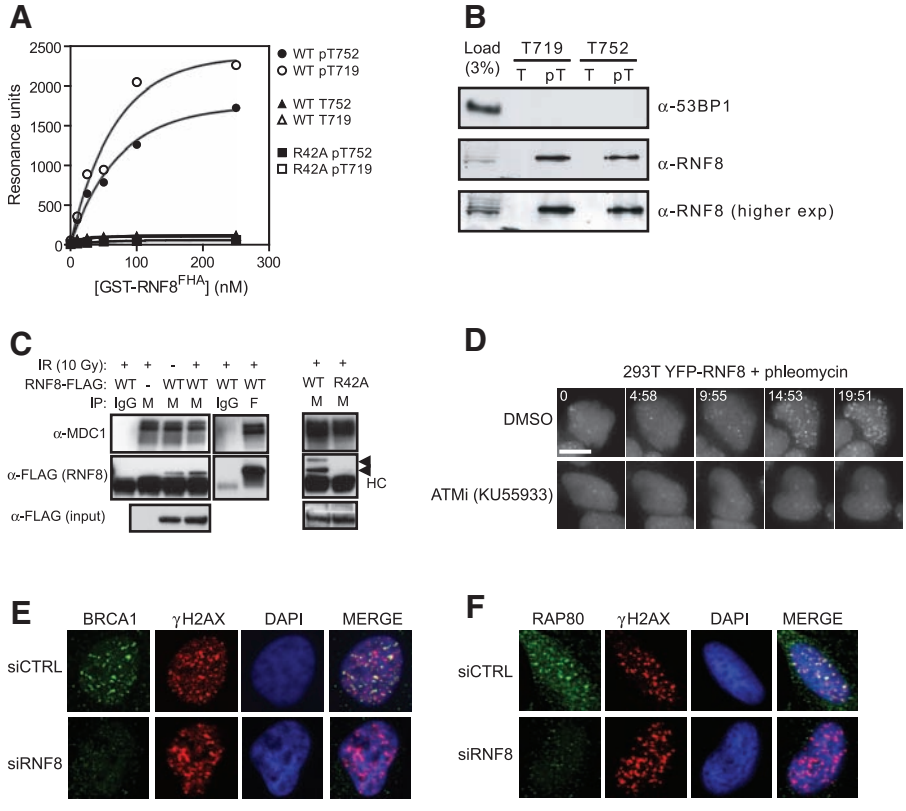


Fig. 3. RNF8 mediates BRCA1-RAP80 IRIF through a physical interaction with MDC1. (**A**) Binding curves of GST-RNF8^{FHA} (WT), the R42A mutant obtained by surface plasmon resonance with peptides corresponding to MDC1 epitopes surrounding T719, T752, or their phosphorylated counterparts (pT719 and pT752). (**B**) Peptide pull-downs of HeLa nuclear extracts with immobilized peptides; phosphorylated (pT) or unphosphorylated (T), encompassing MDC1 T719 or T752 residues. (**C**) RNF8 interacts with MDC1 in vivo. Extracts from 293T cells mock-transfected (–) or transfected (+) with RNF8-FLAG (WT) or the R42A FHA mutant were immunoprecipitated (IP) with antibodies to MDC1 (M), FLAG (F), or normal mouse immunoglobulin G (IgG) and probed for MDC1 or RNF8-FLAG as indicated. Arrowheads indicate the RNF8-specific signal. RNF8 appears modified when interacting with MDC1. HC, IgG heavy chains. (**D**) Time-lapse microscopy of 293T cells stably expressing YFP-RNF8 preincubated with ATM inhibitor KU55933 or dimethyl sulfoxide and treated with the radio-mimetic drug phleomycin (1.5 mg/ml) for the indicated times (min:sec). Three-dimensional (3D) image data sets were computationally deconvolved and shown as 2D projections. Scale bar, 10 μm. (**E** and **F**) Irradiated (10 Gy) HeLa cells transfected with the indicated siRNAs were stained with antibodies to γH2AX, BRCA1 (D), or RAP80 (E) 1 hour after IR.

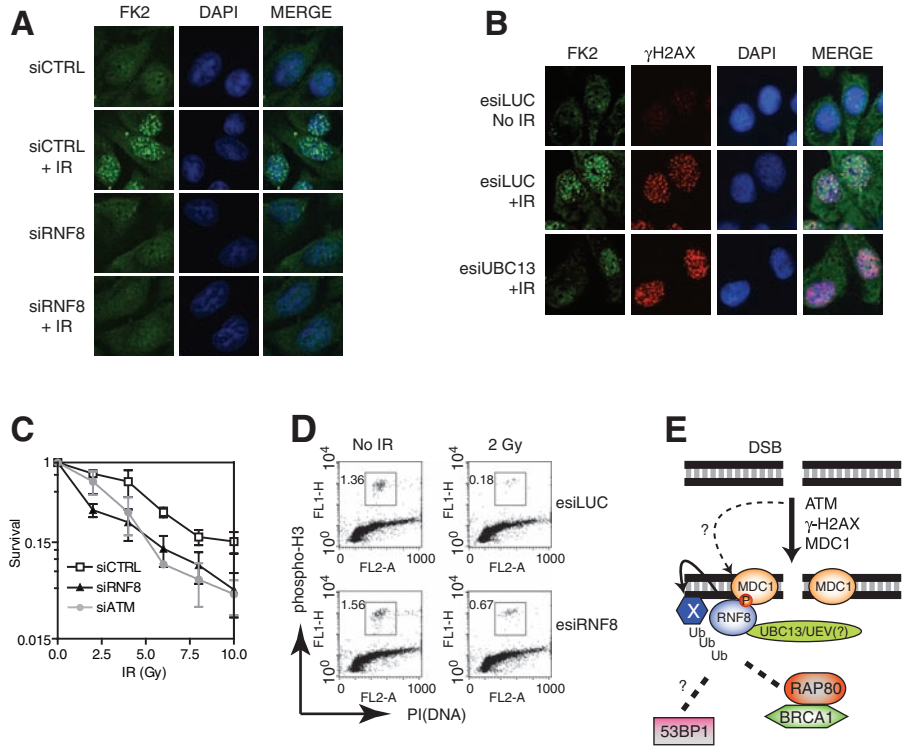


Fig. 4. RNF8 cooperates with UBC13 to mediate ubiquitin IRIF and a functional DDR. (**A** and **B**) Irradiated (10 Gy) U2OS cells transfected with the indicated siRNA (A) or esiRNAs (B) were stained with FK2 antibodies to conjugated ubiquitin and γH2AX to assess IRIF. Cells were fixed 1 hour after IR. (**C**) Clonogenic survival of HeLa cells transfected with siRNAs against ATM (siATM), RNF8 (siRNF8), or a non-targeting control (siCTRL). $N = 3 \pm$ SEM. (**D**) G_2/M checkpoint analysis of U2OS cells transfected with the indicated esiRNAs. Fixed mock-treated (No IR) or irradiated (2 Gy) cells were stained with an antibody to phospho-histone H3 and propidium iodide (PI). The percentage of mitotic cells was determined by fluorescence-activated cell sorting. (**E**) Model of RNF8 action at DSBs. RNF8 is recruited by ATM-phosphorylated MDC1 to DSBs, where it ubiquitinates an unknown protein (X), recruiting RAP80-BRCA1 and allowing 53BP1 to recognize methylated histones, as suggested by the RNF8-dependent recruitment of the 53BP1 Tudor domain to DSB sites (fig. S13). Our results also suggest the presence of a MDC1-independent pathway of RNF8 action (dashed line) that mediates 53BP1 and BRCA1 IRIF formation.

mediate IRIF formation by RAP80 and conjugated ubiquitin. Indeed, RNF8 was essential for IRIF formation in both cases (Figs. 3F and 4A, and fig. S10A). Because RAP80 is not required for 53BP1 focus formation after IR (*14*) and 53BP1 is not needed for RAP80 IRIF (fig. S10B), these results indicated that RNF8 acts downstream of MDC1 to promote at least two types of IRIF: those containing 53BP1 and those containing BRCA1. In support of this model, mutation of the TQXF motifs in MDC1 also impaired BRCA1 and conjugated ubiquitin IRIF (fig. S11).

RNF8 can bind to multiple E2-conjugating enzymes to catalyze both K63-linked and K48-linked ubiquitin chains (*23*, *25*). When we screened a panel of 13 E2 enzymes by RNAi and quantitative microscopy, only UBC13 depletion markedly impaired 53BP1 IRIF formation (fig. S12). UBC13 depletion also impaired IRIF by conjugated ubiquitin (Fig. 4B). Entirely consistent with our data, genetic ablation of UBC13 in DT40 cells abrogate BRCA1 and ubiquitin IRIF (*26*). As UBC13 is the only known E2-conjugating enzyme that catalyzes K63-linked polyubiquitination, these findings are consistent with data indicating that ubiquitin IRIF form in part through K63-linked polyubiquitination (*15*, *26*). Moreover, as UBC13 physically interacts with RNF8 to catalyze K63-linked ubiquitin chains (*27*), the available data imply that RNF8 is responsible for formation of K63-linked ubiquitin chains at DSB sites.

Consistent with RNF8 playing an important function in the DDR, we found that RNF8 depletion caused IR hypersensitivity in clonogenic cell-survival assays (Fig. 4C). Furthermore, a greater proportion of RNF8-depleted cells progressed into M phase after irradiation than did cells transfected with a control esiRNA, which indicates that RNF8 enforces the G_2/M DNA-damage checkpoint (Fig. 4D).

Our results identify mammalian RNF8 as an important component of the DDR. Specifically, RNF8 binds to ATM-target motifs on MDC1, thus recruiting RNF8 to DSB sites. RNF8 then triggers the formation of ubiquitin conjugates that promote recruitment of the RAP80-BRCA1 complex and 53BP1 to DSB sites, thereby enhancing DNA-damage checkpoint events and promoting cell survival (Fig. 4E).

References and Notes

1. J. Bartek, J. Lukas, *Curr. Opin. Cell Biol.* **19**, 238 (2007).
2. R. S. Maser, K. J. Monsen, B. E. Nelms, J. H. Petrini, *Mol. Cell. Biol.* **17**, 6087 (1997).
3. T. Stiff *et al.*, *Cancer Res.* **64**, 2390 (2004).
4. T. T. Paull *et al.*, *Curr. Biol.* **10**, 886 (2000).
5. A. Celeste *et al.*, *Nat. Cell Biol.* **5**, 675 (2003).
6. S. Bekker-Jensen *et al.*, *J. Cell Biol.* **173**, 195 (2006).
7. M. Stucki *et al.*, *Cell* **123**, 1213 (2005).
8. Z. Lou *et al.*, *Mol. Cell* **21**, 187 (2006).
9. M. S. Lee, R. A. Edwards, G. L. Thede, J. N. Glover, *J. Biol. Chem.* **280**, 32053 (2005).
10. S. Bekker-Jensen, C. Lukas, F. Melander, J. Bartek, J. Lukas, *J. Cell Biol.* **170**, 201 (2005).
11. G. S. Stewart, B. Wang, C. R. Bignell, A. M. Taylor, S. J. Elledge, *Nature* **421**, 961 (2003).
12. M. Goldberg *et al.*, *Nature* **421**, 952 (2003).
13. J. Yan *et al.*, *Cancer Res.* **67**, 6647 (2007).
14. H. Kim, J. Chen, X. Yu, *Science* **316**, 1202 (2007).
15. B. Sobhian *et al.*, *Science* **316**, 1198 (2007).
16. B. Wang *et al.*, *Science* **316**, 1194 (2007).
17. S. Matsuoka *et al.*, *Science* **316**, 1160 (2007).
18. R. A. DiTullio Jr. *et al.*, *Nat. Cell Biol.* **4**, 998 (2002).
19. L. B. Schultz, N. H. Chehab, A. Malikzay, T. D. Halazonetis, *J. Cell Biol.* **151**, 1381 (2000).
20. V. Plans, M. Guerra-Rebollo, T. M. Thomson, *Oncogene*, in press. Published online 3 September 2007 (10.1038/sj.onc.1210782).
21. R. Kittler, A. K. Heninger, K. Franke, B. Habermann, F. Buchholz, *Nat. Methods* **2**, 779 (2005).
22. D. Durocher, J. Henckel, A. R. Fersht, S. P. Jackson, *Mol. Cell* **4**, 387 (1999).
23. K. Ito *et al.*, *Eur. J. Biochem.* **268**, 2725 (2001).
24. I. Hickson *et al.*, *Cancer Res.* **64**, 9152 (2004).
25. J. Bothos, M. K. Summers, M. Venere, D. M. Scolnick, T. D. Halazonetis, *Oncogene* **22**, 7101 (2003).
26. G. Y. Zhao *et al.*, *Mol. Cell* **25**, 663 (2007).
27. V. Plans *et al.*, *J. Cell. Biochem.* **97**, 572 (2006).

28. We thank members of the Durocher and Jackson laboratories for input on the manuscript and M. Vojvodic for experimental assistance. We also thank A. C. Gingras for the stable YFP-RNF8 line, KuDOS Pharmaceuticals for providing inhibitors and reagents, and Abcam for the MDC1 pT719 antibody. This work was supported by grants from the Canadian Institutes of Health Research (CIHR) to D.D. and by funding from Cancer Research UK and the European Union (S.P.J.). N.K.K. is a CIHR postdoctoral fellow and an alumnus of the Excellence in Radiation Research for the 21st Century training program; F.D.S. holds a Terry-Fox studentship from the National Cancer Institute of Canada; S.N. is a Gail-Posluns Fellow and is supported by the Mitsubishi Pharma Research Foundation. D.D. is a Canada Research Chair (Tier II) in Proteomics, Functional Genomics, and Bioinformatics.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1150034/DC1
Materials and Methods
Figs. S1 to S14
References

4 September 2007; accepted 5 November 2007
Published online 15 November 2007;
10.1126/science.1150034
Include this information when citing this paper.

Engineering Modified Bt Toxins to Counter Insect Resistance

Mario Soberón,[1]* Liliana Pardo-López,[1] Idalia López,[1] Isabel Gómez,[1] Bruce E. Tabashnik,[2] Alejandra Bravo[1]*

The evolution of insect resistance threatens the effectiveness of *Bacillus thuringiensis* (Bt) toxins that are widely used in sprays and transgenic crops. Resistance to Bt toxins in some insects is linked with mutations that disrupt a toxin-binding cadherin protein. We show that susceptibility to the Bt toxin Cry1Ab was reduced by cadherin gene silencing with RNA interference in *Manduca sexta*, confirming cadherin's role in Bt toxicity. Native Cry1A toxins required cadherin to form oligomers, but modified Cry1A toxins lacking one α-helix did not. The modified toxins killed cadherin-silenced *M. sexta* and Bt-resistant *Pectinophora gossypiella* that had cadherin deletion mutations. Our findings suggest that cadherin promotes Bt toxicity by facilitating toxin oligomerization and demonstrate that the modified Bt toxins may be useful against pests resistant to standard Bt toxins.

The toxins produced by *Bacillus thuringiensis* (Bt) kill some major insect pests such as mosquitoes and crop-eating caterpillars but are harmless to vertebrates and most other organisms (*1*). Transgenic corn and cotton producing Bt toxins grew on more than 32 million hectares worldwide in 2006 (*2*). The primary threat to the long-term efficacy of Bt toxins is the evolution of resistance by pests (*3*–*5*). Many insects have been selected for resistance to Bt toxins in the laboratory, and two crop pests (*Plutella xylostella* and *Trichoplusia ni*) have evolved resistance to Bt sprays outside of the laboratory (*3*–*13*). The most widely used Bt toxins are crystal toxins in the Cry1A family, particularly Cry1Ab in transgenic Bt corn and Cry1Ac in transgenic Bt cotton, which kill caterpillars (lepidopteran larvae) (*7*). Cry1A toxins bind to the extracellular domain of cadherin proteins that traverse the insect larval midgut membrane (*14*). Disruption of Bt toxin binding to midgut receptors is the most common mechanism of insect resistance (*6*). Mutations in the midgut cadherins that bind Cry1Ac are linked with and probably cause resistance in at least three lepidopteran pests of cotton (*5*, *10*, *12*).

Two hypotheses proposed to explain how Cry1A toxins function are the pore-formation model and the signaling model (*15*, *16*). These theories share initial steps: Cry1A protoxins are ingested, solubilized in the gut, and cleaved by midgut proteases such as trypsin to yield activated 60-kD monomeric toxins that bind to cadherin with high affinity (*14*). The signaling model, derived from studies of insect cell cultures, suggests that after protease-activated monomeric toxins bind to cadherin, initiation of a magnesium-dependent signaling pathway causes cell death (*16*). In contrast, on the basis of results from in vitro experiments and bioassays, the pore-formation model proposes that protease-activated monomers bind to cadherin to facilitate protease cleavage of the N terminus of the toxin, including helix α-1of domain I (*17*, *18*). This cleavage induces the assembly of oligomeric forms of the toxin, which have increased binding affinity to secondary receptors,

[1]Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, Cuernavaca 62250, Morelos, Mexico. [2]Department of Entomology, University of Arizona, Tucson, AZ, USA.

*To whom correspondence should be addressed. E-mail: bravo@ibt.unam.mx (A.B.); mario@ibt.unam.mx (M.S.)

including glycosylphosphatidylinositol-anchored proteins, aminopeptidase N, and alkaline phosphatases (*15*, *19*). After oligomers bind to secondary receptors, they create pores in the midgut membrane that cause osmotic shock and cell death (*15*).

According to the pore-formation model, the binding of protease-activated toxin to cadherin is essential for the removal of helix α-1, which in turn promotes oligomerization. Therefore, we hypothesized that modified Cry1Ab and Cry1Ac toxins lacking helix α-1 (referred to hereafter as Cry1AbMod and Cry1AcMod) could form oligomers without cadherin. To test this hypothesis, we compared oligomerization of native and modified Cry1Ab and Cry1Ac in the presence and absence of toxin-binding cadherin fragments. Toxins were activated by trypsin to mimic the protease activation that occurs in the insect gut. Previous work shows that trypsin-activated Cry1Ab forms 250-kD oligomers in the presence of *Manduca sexta* cadherin fragments containing toxin-binding regions corresponding to cadherin repeats 7 and 11 or a single-chain antibody (scFv73) that mimics these cadherin toxin-binding regions (*17*, *20*). We found that trypsin-activated Cry1Ab and Cry1Ac formed oligomers in the presence of a protein fragment corresponding to *M. sexta* cadherin repeat 12 (CADR12), an important toxin-binding region (*21*), but not without cadherin or with cadherin repeat 9 (CADR9), a region that does not bind toxin (*21*) (Fig. 1). In contrast, trypsin-activated Cry1AbMod and Cry1AcMod formed oligomers without cadherin, although not as efficiently as the wild type with CADR12 (Fig. 1).

We used RNA interference (RNAi) to reduce production of cadherin protein (Fig. 2 and fig. S1) and decrease the susceptibility of *M. sexta* larvae to Cry1Ab. Larvae injected with either 1 μg of cadherin double-stranded RNA (dsRNA) or water only (control) ate a diet treated with 20 ng of Cry1Ab protoxin/cm^2. After 3 days, survival was 92% for 48 RNAi-treated larvae versus 0% for 48 control larvae (chi-square test, $P < 0.001$).

The next experiment showed that cadherin-silenced *M. sexta* larvae were much more susceptible to Cry1AbMod than to Cry1Ab (fig. S2). Confirming the results described above, the survival of larvae fed a diet treated with 20 ng of Cry1Ab protoxin/cm^2 was higher for 48 RNAi-treated larvae (92%) than for 48 control larvae (2%) (chi-square test, $P < 0.001$). However, on a diet with 5 ng of Cry1AbMod protoxin/cm^2, 48 RNAi-treated larvae had only 2% survival, which is significantly lower than the aforementioned 92% survival of 48 RNAi-treated larvae fed a diet treated with 20 ng of Cry1Ab protoxin/cm^2 (chi-square test, $P < 0.001$).

We also tested modified and native Cry1A toxins against larvae of a resistant strain (AZP-R) and a susceptible strain (APHIS-S) of the cotton

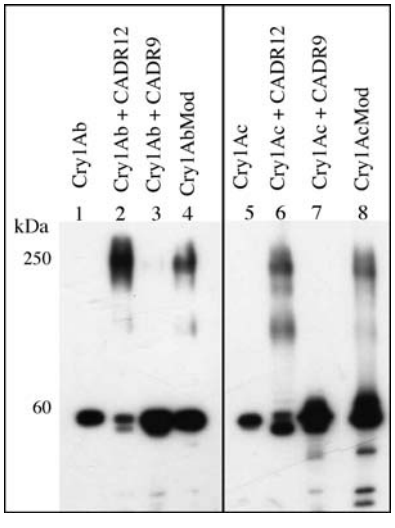


Fig. 1. Oligomer formation by native and modified Bt toxins. Cadherin fragments added correspond to regions that bind toxin (CADR12, lanes 2 and 6) and do not bind toxin (CADR9, lanes 3 and 7). Western blots were probed with polyclonal antibodies to Cry1Ab (lanes 1 to 4) or to Cry1Ac (lanes 5 to 8).

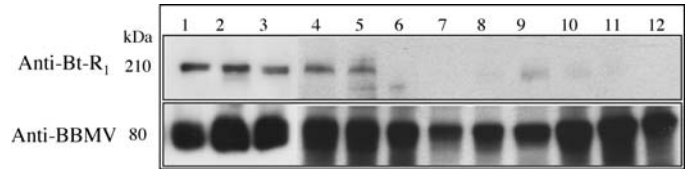


Fig. 2. Effect of RNAi on cadherin protein expression in *M. sexta* larvae. Western blots were tested for the *M. sexta* cadherin protein (BT-R_1) and for an 80-kD brush-border membrane vescicle protein (BBMV). Lanes 1 to 5: Control larvae injected with water only and fed a diet without toxin. Lanes 6 to 12: Larvae injected with 1 μg of BT-R_1 dsRNA and fed a diet with 20 ng of Cry1Ab/cm^2.

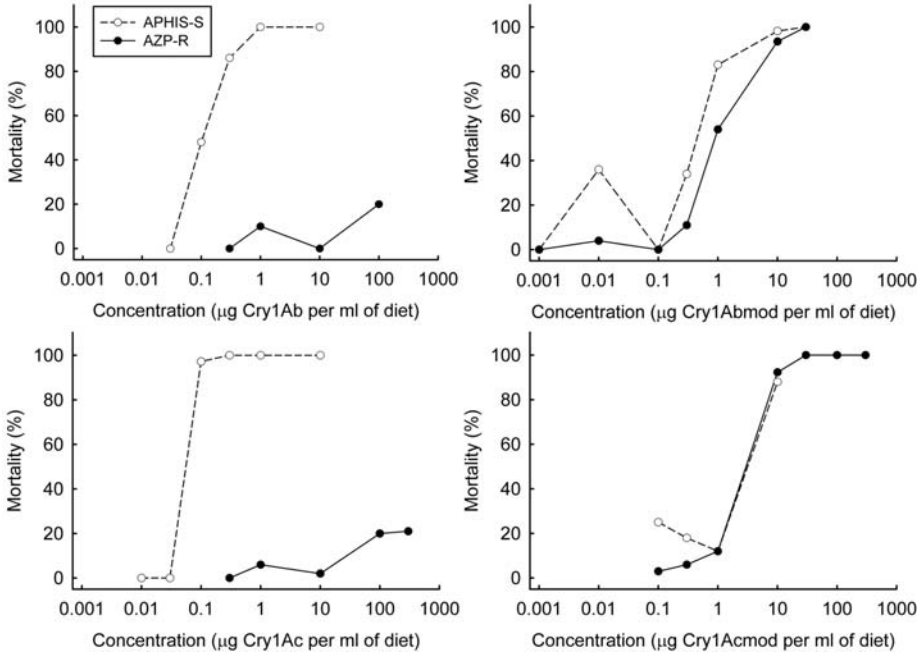


Fig. 3. Responses of susceptible (APHIS-S) and resistant (AZP-R) pink bollworm larvae to native toxins [Cry1Ab and Cry1Ac (**left**)] and modified toxins [Cry1AbMod and Cry1AcMod (**right**)].

Table 1. Modified toxins Cry1AbMod and Cry1AcMod kill pink bollworm larvae resistant to Cry1Ab and Cry1Ac.

Toxin	Insect strain[a]	*n*	LC_{50} (95% FL)[b]	Resistance ratio[c]
Cry1Ab	Resistant	320	>100[d]	>910
	Susceptible	320	0.11 (0.09 – 0.14)	
Cry1AbMod	Resistant	400	1.1 (0.73 – 1.6)	2.8
	Susceptible	400	0.39 (0.28 – 0.51)	
Cry1Ac	Resistant	340	>300[e]	>3700
	Susceptible	400	0.079	
Cry1AcMod	Resistant	460	2.8 (1.8 – 3.8)	0.41
	Susceptible	320	6.8	

[a]The resistant strain was AZP-R; the susceptible strain was APHIS-S. [b]LC_{50} of larvae in micrograms of protoxin per milliliter of diet; 95% fiducial limits (FL) are shown in parentheses when available. [c]LC_{50} of resistant strain divided by LC_{50} of susceptible strain. [d]The highest concentration tested (100 μg of protoxin/ml of diet) killed only 20% of larvae. [e]The highest concentration tested (300 μg of protoxin/ml of diet) killed only 21% of larvae.

pest pink bollworm (*Pectinophora gossypiella*). Larvae from AZP-R, but not APHIS-S, survived on Bt cotton producing Cry1Ac (*8*, *9*). This resistance to Cry1Ac in AZP-R is linked with deletion mutations in the cadherin receptor gene (*10*). Cry1AbMod and Cry1AcMod reduced or overcame resistance in the AZP-R strain (Fig. 3 and Table 1). On the basis of the concentration that killed 50% (LC_{50}) for AZP-R relative to APHIS-S, AZP-R was >910-fold resistant to Cry1Ab and >3700-fold resistant to Cry1Ac (Table 1). In contrast, AZP-R was only 2.8-fold resistant to Cry1AbMod and was not resistant to Cry1AcMod (Table 1). Against resistant larvae, the LC_{50} was more than 100 times higher for Cry1Ac than for Cry1AbMod or Cry1AcMod (Table 1). Conversely, against susceptible larvae, the native toxins were more potent than the modified toxins. This implies that, relative to native toxins against susceptible larvae, modified toxins had lower stability in the midgut, reduced oligomer-forming ability (Fig. 1), or reduced ability of oligomers to ultimately cause mortality.

The results suggest that in two species of Lepidoptera, cadherin receptor protein in the larval midgut mediates the toxicity of Cry1A toxins by facilitating removal of helix α-1, which promotes toxin oligomerization. The modified toxins Cry1AbMod and Cry1AcMod lacking helix α-1 formed oligomers in vitro without cadherin, whereas native Cry1Ab and Cry1Ac did not. The modified toxins killed insects with greatly reduced susceptibility to native Cry1A toxins caused by RNAi silencing of the cadherin gene or by mutations in the cadherin gene. These results support the pore-formation model (*15*) and not the signaling model, which does not include removal of helix α-1 or toxin oligomerization (*16*).

If the results seen here with the pink bollworm extend to other lepidopterans, Cry1AbMod and Cry1AcMod could be broadly useful for countering or delaying pest resistance to Cry1A toxins. However, we do not know whether Cry1AMod toxins kill insects with mechanisms of resistance unrelated to cadherin, such as the disruption of other receptors or decreased protease activation (*22*). Many Bt toxins have structural topology similar to Cry1A, form oligomers, and induce pores (*17*, *23*–*28*), suggesting that they share a similar mode of action. It remains to be determined whether, parallel to results with Cry1A toxins, other Cry toxins lacking helix α-1 can kill resistant insects that have altered receptors. In addition, insects can probably evolve resistance to modified Bt toxins lacking helix α-1. Nonetheless, along with native Bt toxins such as Cry2 and Vip3 (*29*) that have not been used as extensively as Cry1A toxins, the modified toxins broaden the options for pest control.

References and Notes

1. R. A. de Maagd, A. Bravo, N. Crickmore, *Trends Genet.* **17**, 193 (2001).
2. C. James, *Int. Serv. Acquis. Agro-Biotech Appl. Briefs* **35**, 1 (2006).
3. B. E. Tabashnik, *Annu. Rev. Entomol.* **39**, 47 (1994).
4. F. Gould, *Annu. Rev. Entomol.* **43**, 701 (1998).
5. L. J. Gahan, F. Gould, D. G. Heckel, *Science* **293**, 857 (2001).
6. J. Ferré, J. Van Rie, *Annu. Rev. Entomol.* **47**, 501 (2002).
7. B. E. Tabashnik *et al.*, *J. Econ. Entomol.* **96**, 1031 (2003).
8. B. E. Tabashnik, T. J. Dennehy, Y. Carrière, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15389 (2005).
9. B. E. Tabashnik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12980 (2000).
10. S. Morin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5004 (2003).
11. B. E. Tabashnik *et al.*, *J. Econ. Entomol.* **95**, 1018 (2002).
12. X. Xu, L. Yu, Y. Wu, *Appl. Environ. Microbiol.* **71**, 948 (2005).
13. A. F. Janmaat, J. H. Myers, *Proc. R. Soc. London Ser. B* **270**, 2263 (2003).
14. R. K. Vadlamudi, E. Weber, I. Ji, T. H. Ji, L. A. Bulla Jr., *J. Biol. Chem.* **270**, 5490 (1995).
15. A. Bravo *et al.*, *Biochim. Biophys. Acta* **1667**, 38 (2004).
16. X. Zhang, M. Candas, N. B. Griko, R. Taissing, L. A. Bulla Jr., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9897 (2006).
17. I. Gómez, J. Sánchez, R. Miranda, A. Bravo, M. Soberón, *FEBS Lett.* **513**, 242 (2002).
18. N. Jiménez-Juárez *et al.*, *J. Biol. Chem.* **282**, 21222 (2007).
19. J. L. Jurat-Fuentes, M. J. Adang, *Eur. J. Biochem.* **271**, 3127 (2004).
20. I. Gómez, D. H. Dean, A. Bravo, M. Soberón, *Biochemistry* **42**, 10482 (2003).
21. G. Hua, J. L. Jurat-Fuentes, M. J. Adang, *J. Biol. Chem.* **279**, 28051 (2004).
22. D. G. Heckel *et al.*, *J. Invertebr. Pathol.* **95**, 192 (2007).
23. A. I. Aronson, C. Glen, L. Wu, *Appl. Environ. Microbiol.* **65**, 2503 (1999).
24. C. Rausell *et al.*, *Biochim. Biophys. Acta* **1660**, 99 (2004).
25. N. J. Tigue, J. Jacoby, D. J. Ellar, *Appl. Environ. Microbiol.* **67**, 5715 (2001).
26. C. Muñoz-Garay *et al.*, *J. Membr. Biol.* **212**, 61 (2006).
27. S. Herrero, J. Gonzalez-Cabrera, J. Ferré, P. L. Bakker, R. A. de Maagd, *Biochem. J.* **384**, 507 (2004).
28. S. Likitvivatanavong, G. Katzenmeier, Ch. Angsuthanasombat, *Arch. Biochem. Biophys.* **445**, 46 (2006).
29. R. E. Jackson, M. A. Marcus, F. Gould, J. R. Bradley Jr., J. W. Duyn, *J. Econ. Entomol.* **100**, 180 (2007).

30. We thank L. Cabrera, C. Unnithan, and R. Biggs for technical assistance in rearing and bioassays; T. J. Dennehy and the Extension Arthropod Resistance Management Laboratory staff for permitting use of the resistant pink bollworm strain they isolated from Arizona cotton fields; A. Gassmann for comments and graphical assistance; and L. Masson, X. Li, and M. Trejo for comments. This work was supported by CONACyT grants 48631-Q, 46176-Q, J46B29-Q, and J45B63-Q; NIH grant 1R01 AI066014; the National Research Initiative; the Cooperative State Research, Education, and Extension Service; and U.S. Department of Agriculture grants 2007-35607-17780, 2003-01469, and 2006-35302-17365.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1146453/DC1
Materials and Methods
Figs. S1 to S4
References

13 June 2007; accepted 19 September 2007
Published online 1 November 2007;
10.1126/science.1146453
Include this information when citing this paper.

Genetically Determined Differences in Learning from Errors

Tilmann A. Klein,[1]* Jane Neumann,[1] Martin Reuter,[2] Jürgen Hennig,[3] D. Yves von Cramon,[1,4] Markus Ullsperger[1,4]*

The role of dopamine in monitoring negative action outcomes and feedback-based learning was tested in a neuroimaging study in humans grouped according to the dopamine D2 receptor gene polymorphism DRD2-TAQ-IA. In a probabilistic learning task, A1-allele carriers with reduced dopamine D2 receptor densities learned to avoid actions with negative consequences less efficiently. Their posterior medial frontal cortex (pMFC), involved in feedback monitoring, responded less to negative feedback than others' did. Dynamically changing interactions between pMFC and hippocampus found to underlie feedback-based learning were reduced in A1-allele carriers. This demonstrates that learning from errors requires dopaminergic signaling. Dopamine D2 receptor reduction seems to decrease sensitivity to negative action consequences, which may explain an increased risk of developing addictive behaviors in A1-allele carriers.

"You learn from your mistakes," people say. We usually learn from both positive and negative action outcomes, which induce reinforcement of successful and avoidance of erroneous behavior, respectively (*1*). The relative amount of learning from successes and errors varies across individuals as a result of disease or pharmacological intervention (*2*). Can even our genetic makeup influence the way we learn from errors? An important factor in the use of negative and positive feedback for learning seems to be the neurotransmitter dopamine (*3*–*5*). A human genetic polymorphism (DRD2-TAQ-IA) is known to modulate dopamine D2 receptor density. The A1 allele is associated with a reduction in D2 receptor density by up to 30% (*6*–*8*). This reduction has been linked to multiple addictive and compulsive behaviors (*9*, *10*), which suggests some insensitivity to negative consequences of self-destructive behavior. This might be linked to

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [2]University of Bonn, Bonn, Germany. [3]University of Giessen, Giessen, Germany. [4]Max Planck Institute for Neurological Research, Cologne, Germany.

*To whom correspondence should be addressed. E-mail: tklein@cbs.mpg.de (T.A.K.); ullsperger@nf.mpg.de (M.U.)

a general deficit in learning from errors. Here, we report patterns of brain activity underlying a reduced ability to use negative feedback for avoidance learning in carriers of the A1 allele. Our findings suggest a genetically driven change in the dynamic interaction of performance monitoring and long-term memory formation. When action outcomes call for adaptations, a performance-monitoring system in the posterior medial frontal cortex (pMFC) signals the need for adjustments (*11*, *12*). The rostral cingulate zone (RCZ), located in the pMFC, has been suggested to be involved in learning from errors (*13*, *14*). A neurobiological theory holds that this region receives dopaminergic teaching signals from the midbrain coding whether an event is better or worse than predicted (*14*). In close interaction with the performance-monitoring system, the basal ganglia, in particular the nucleus accumbens (NAC), play a major role in reward-based learning (*12*, *15*–*17*). Moreover, the performance-monitoring system needs to interact with the hippocampal formation to enable learning of stimulus-reward associations.

To investigate neural activity related to error-based learning, we recorded functional magnetic resonance imaging (fMRI) data from 26 healthy male subjects grouped by genotype [A1-allele carrier, A1+ group, $n = 12$; non–A1-allele carrier, A1– group, $n = 14$ (*18*)]. We used a probabilistic learning task sensitive to dopaminergic manipulations (*2*). Participants had to learn to choose the more-often rewarded symbols from pairs of stimuli presented in random order. After each choice, probabilistic feedback was provided (Fig. 1, top). After learning, participants were confronted in a behavioral posttest with the same symbols, now paired with symbols other than the one from the learning phase [supporting online material (SOM), table S1]. This allowed us to disentangle preference for the most-often rewarded symbol "A" and avoidance of the least-often rewarded symbol "B."

The groups defined by the presence or the absence of the A1 allele did not differ in the average frequency of selecting favorable symbols nor in the rate of negative feedback; however, we found a remarkable group difference in avoidance learning (Fig. 1, bottom left) (SOM text). In the posttest, the A1+ group avoided the negative symbol B significantly less than they chose the positive symbol A ($P = 0.03$). Moreover, their avoidance of B was reduced compared with the A1– group ($P = 0.03$), who did not show a significant difference between selecting A and avoiding symbol B ($P = 0.17$). Consistent with this behavior, they also showed a reduced negative feedback–related fMRI signal in the RCZ ($x = 4$, $y = 24$, $z = 33$; z score = 3.5, 324 mm^3) compared with the A1– group (Fig. 2A, and table S2). In a Bayesian analysis (*18*, *19*), we observed a posterior probability of 95.8% for a group difference in RCZ activity induced by negative feedback. Moreover, only members of the A1– group showed positive correlations with negative feedback–related RCZ activity and preference for the A symbol ($r = 0.53$, $P = 0.05$) and avoidance of the B symbol ($r = 0.55$, $P = 0.04$). A further strong signal increase on negative feedback in the right middorsal prefrontal cortex [$x = 40$, $y = 21$, $z = 27$; z score = 4.3, middle frontal gyrus (MFG)] was found only in the A1– group (posterior probability of group difference: 97.1%).

To study learning over the course of the probabilistic learning task, we modeled subjects' behavior using a modified Rescorla-Wagner reinforcement learning model (*20*) (fig. S1). In this computational model, the difference in activity of the output neurons provides a trial-by-trial estimate of certainty of the given response. The A1– group reached a significantly higher response certainty in the last third ($t = 2.2$, $P = 0.04$). The development of the certainty over the course of the experiment is shown in Fig. 1 (bottom right). In both groups, the curves resemble a logarithmic learning curve with a steep increase in the first third and an asymptotic course at the end of the experiment. After an initial period of about 200 trials, the A1– group developed a higher response certainty than the A1+ group. For both genetic groups, response certainty negatively correlated with pMFC activity (fig. S2). Note that, in the A1– group, the time course of certainty, which reflected learning progress, showed a positive correlation with activity in the posterior hippocampus bilaterally ($x = 22$, $y = –39$, $z = 6$; z score = 3.9, 216 mm^3 and $x = –23$, $y = –39$, $z = 3$; z score = 3.5, 81 mm^3), whereas no such correlation was found in A1+ participants [Bayesian posterior probability of group difference: right, 94.9%; left, 96.2% (Fig. 2B) and table S3].

How does feedback monitoring in the RCZ interact with forming memories in the hippocampus? Anatomically, these areas are connected via the cingulate bundle. To investigate learning-related changes in functional interactions of the RCZ and other brain areas over time, we performed a psychophysiological interaction analysis (PPI) (*21*). The experiment was divided into three parts of equal length. We then contrasted the functional connectivity of the RCZ observed in the first third with the connectivity observed in the last third of the learning experiment, thereby capturing the difference between steep rule acquisition in the beginning and more stable rule exploitation at the end. Again, in the A1– group, we observed a significant change over time: In the first third of the experiment, the functional coupling between the RCZ activity and the bilateral hippocampus was substantially stronger

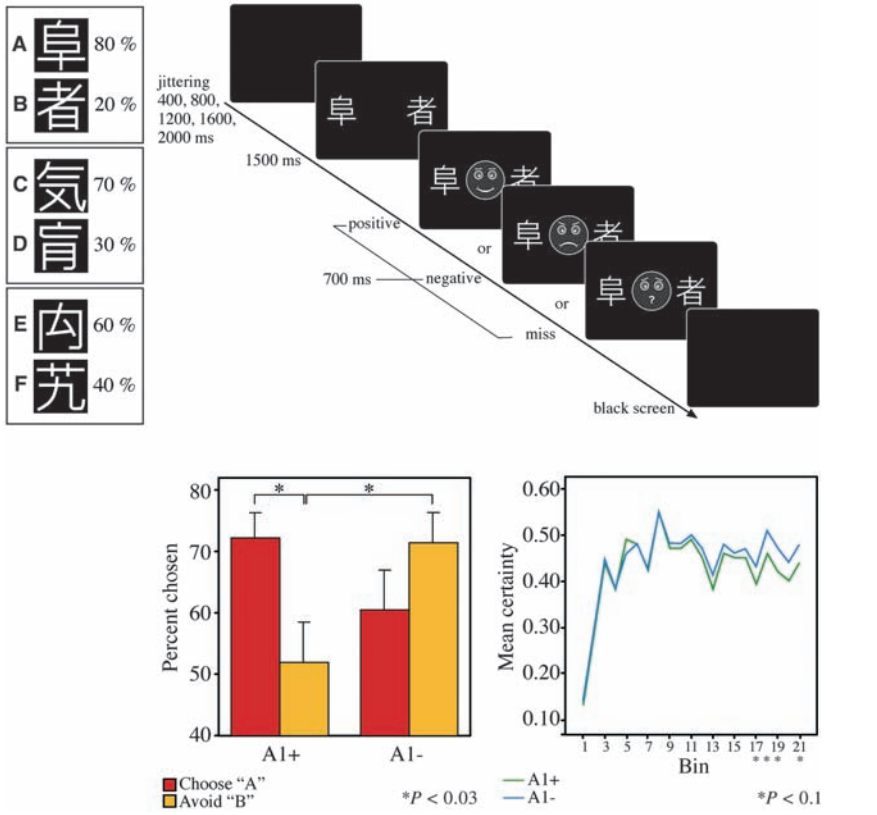


Fig. 1. Probabilistic learning task, behavioral and computational results. (Top) Stimuli, reward probabilities (percent positive feedback), and schematic trial sequence of the probabilistic learning task (*2*). (Bottom, left) Result of the behavioral post test: Choosing the good symbol (A) and avoiding the bad one (B) differs between the two genetic groups (group × selection interaction: $F_{1,24} = 8.1$, $P = 0.009$). (Bottom, right) Certainty of the given response resulting from the computational model, binned into bins of 20 trials each and differentiated between the two genetic groups.

than in the last third (Fig. 2C). The A1+ group showed no such correlation (Bayesian posterior probability of group difference: left hippocampus, 99.98%; right hippocampus, 99.91%). Furthermore, only the A1– group showed a similar change in functional coupling between NAC and RCZ over the time course of the experiment (Bayesian posterior probability: 99.54%). The NAC, another major target of dopaminergic projections, has also been implicated in feedback-based decision-making (*12*, *22*–*24*). The fMRI signal in the NAC on both sides was increased by positive feedback as compared with negative feedback (Fig. 2A). This reward-related activity increase was reduced in the A1+ group in the right NAC ($x = 16$, $y = 9$, $z = –6$; z score = –3.96; Bayesian posterior probability of group difference: 94.8%; on the left side, posterior probability reached only 74.1%).

Taken together, our results confirm that dopamine plays a major role in performance monitoring and behavioral modification for reaching optimal performance levels: Alterations in dopaminergic transmission lead to corresponding alterations in negative feedback processing and, related to this, to differences in learning from negative feedback. It appears that reduced dopamine D2 receptor density is associated with reduced capacity to learn negative characteristics of a stimulus from negative feedback. High receptor density in the A1– group is associated with clear avoidance of the most negative stimulus, whereas a reduced receptor density in A1+ subjects is not. Corresponding to this, subjects with a reduced receptor density show a weaker blood oxygen level–dependent (BOLD) response to negative feedback in the performance-monitoring network consisting of the pMFC and basal ganglia. In the pMFC, this difference was specific to negative feedback; its response to positive feedback and negative correlation with certainty (*12*) did not differ between groups (Fig. 2A). Negative feedback–related pMFC activity predicted posttest performance in the A1– group, which suggests that they used negative feedback for avoidance, as well as preference learning. Interestingly, anterior insular activity, thought to be involved in autonomic responses to errors (*12*), was present in both groups, which suggests that the genotype-effect is specific to learning from errors. The differential activity in the MFG, a brain region commonly found in working memory tasks (*25*), may suggest that A1– participants used a monitoring-within-memory strategy of keeping track with selection outcome history. This speculation is supported by the role of prefrontal D2 receptors in working memory functions (*26*, *27*).

Hence, the genetically driven differences in avoidance learning seem to result from a weaker neuronal response to negative feedback. Reduced monitoring signals are less likely to influence the memory system. This is supported by the finding of a reduced interaction of performance monitoring in pMFC and memory-formation in the hippocampus.

It is noteworthy that the fMRI signal reduction in the A1– group is specific to performance monitoring–related processes. It does not generalize to other task-specific activity (SOM text and fig. S4).

At first sight, our findings that subjects with lower D2-receptor densities show reduced avoidance learning may appear to conflict with results indicating that patients with Parkinson's disease on medication, i.e., with enhanced dopaminergic transmission, have problems in learning the negative value of stimuli (*2*). This apparent discrepancy can be resolved by a recent study, which revealed a higher rate of dopamine synthesis in the striatum for subjects with the A1+ configuration compared with A1– subjects (*28*). A reduction in D2 receptors could also affect D2 autoreceptors, which in turn leads to a higher synthesis rate of dopamine. Accordingly, transmission via the unaffected D1 receptors should be strengthened, whereas modulation of phasic postsynaptic D2 activity should be relatively reduced. This should lead to a relative decrease in avoidance learning and a shift to learning mainly from positive reinforcement (*2*, *5*). Parkinson's disease is often treated with tonically acting direct D2 agonists, which also reduce phasic modulations at postsynaptic D2 receptors. A phasic decrease in dopamine, as is suggested to occur after negative feedback (*4*, *14*), may thus be less effective in both studies. This dulled D2-mediated dopaminergic signal in turn would finally lead to a weaker hemodynamic response in the RCZ.

Many studies have found relations between a reduced dopamine D2 receptor density and addiction, obesity, or compulsive gambling (*9*, *10*, *29*). It may be speculated that the insensitivity to negative consequences of an action, as described above, is one feature of a low D2

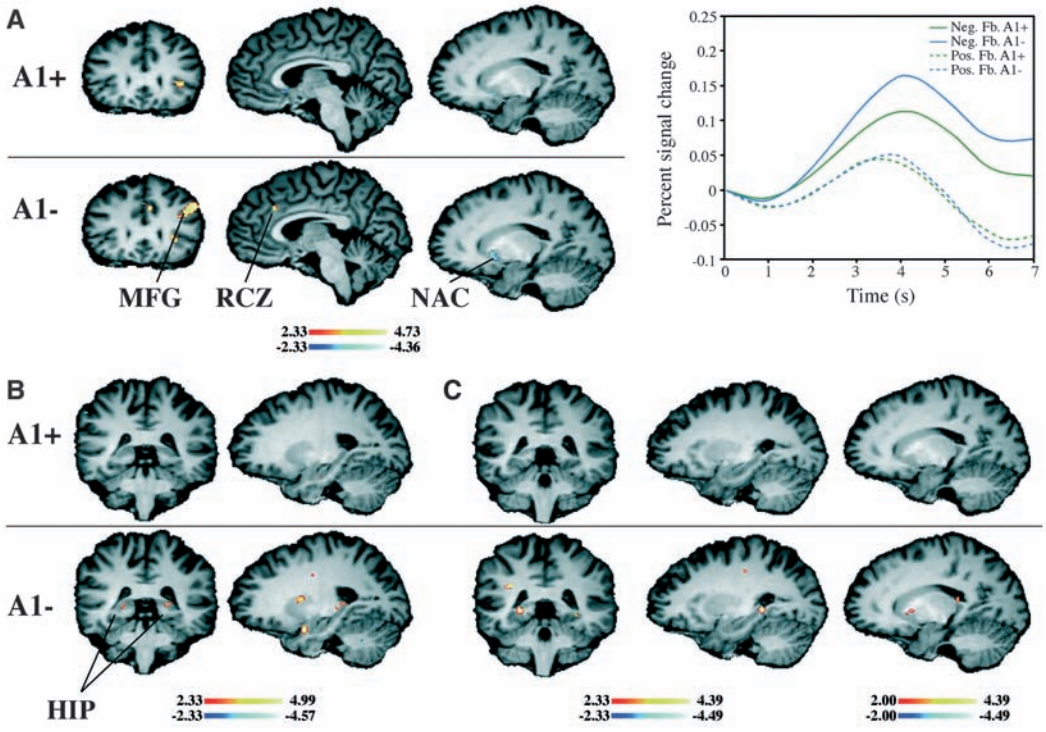


Fig. 2. Genetic influences on the fMRI results. Only clusters with at least 81 mm^3 activated at $z \geq 3.09$ are shown. For visualization, the map thresholds are set at $z = 2.33$ (unless stated otherwise). (**A**) (Left) The contrast between negative and positive feedback for the two genetic groups is shown projected onto a coronal slice ($y = 24$) and two sagittal slices ($x = 4$ and $x = 16$); red, negative feedback > positive feedback; blue, positive feedback > negative feedback. (Right) Percent signal change for positive and negative feedback taken from RCZ ($x = 4$, $y = 24$, $z = 33$). (**B**) Parametric within subject fMRI analysis using the certainty of the given response as a regressor, projected onto a coronal ($y = –39$) and a sagittal ($x = 22$) slice. HIP, hippocampus. (**C**) Psychophysiological interaction analysis between RCZ ($x = 4$, $y = 24$, $z = 33$) and other brain areas, projected onto a coronal ($y = –42$) and two sagittal ($x = –26$ and $x = 16$) slices. Red, stronger interaction in the first third than in the last third of the experiment; blue, stronger interaction in the last than in the first third.

receptor configuration and promotes behavior that could threaten health or social interactions.

References and Notes

1. E. L. Thorndike, *Animal Intelligence: Experimental Studies* (MacMillan, New York, 1911).
2. M. J. Frank, L. C. Seeberger, R. C. O'Reilly, *Science* **306**, 1940 (2004).
3. W. Schultz, *J. Neurophysiol.* **80**, 1 (1998).
4. W. Schultz, *Neuron* **36**, 241 (2002).
5. M. J. Frank, *J. Cogn. Neurosci.* **17**, 51 (2005).
6. T. Ritchie, E. P. Noble, *Neurochem. Res.* **28**, 73 (2003).
7. E. G. Jonsson *et al.*, *Mol. Psychiatry* **4**, 290 (1999).
8. T. Pohjalainen *et al.*, *Mol. Psychiatry* **3**, 256 (1998).
9. E. P. Noble, *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **116**, 103 (2003).
10. A. Bowirrat, M. Oscar-Berman, *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **132**, 29 (2005).
11. K. R. Ridderinkhof, M. Ullsperger, E. A. Crone, S. Nieuwenhuis, *Science* **306**, 443 (2004).
12. M. Ullsperger, D. Y. von Cramon, *J. Neurosci.* **23**, 4308 (2003).
13. R. B. Mars *et al.*, *Neuroimage* **28**, 1007 (2005).
14. C. B. Holroyd, M. G. Coles, *Psychol. Rev.* **109**, 679 (2002).
15. R. Cools, L. Clark, A. M. Owen, T. W. Robbins, *J. Neurosci.* **22**, 4563 (2002).
16. G. Pagnoni, C. F. Zink, P. R. Montague, G. S. Berns, *Nat. Neurosci.* **5**, 97 (2002).
17. M. Ullsperger, D. Y. von Cramon, *J. Cogn. Neurosci.* **18**, 651 (2006).
18. Materials and methods are available as supporting online material on *Science* Online.
19. J. Neumann, G. Lohmann, *Neuroimage* **20**, 1346 (2003).
20. P. F. Rodriguez, A. R. Aron, R. A. Poldrack, *Hum. Brain Mapp.* **27**, 306 (2006).
21. D. R. Gitelman, W. D. Penny, J. Ashburner, K. J. Friston, *Neuroimage* **19**, 200 (2003).
22. B. Knutson, C. M. Adams, G. W. Fong, D. Hommer, *J. Neurosci.* **21**, RC159 (2001).
23. R. Cools, L. Clark, T. W. Robbins, *J. Neurosci.* **24**, 1129 (2004).
24. H. R. Heekeren *et al.*, *Neuroreport* **18**, 951 (2007).
25. M. Petrides, B. Alivisatos, A. C. Evans, E. Meyer, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 873 (1993).
26. M. A. Mehta, R. Swainson, A. D. Ogilvie, J. Sahakian, T. W. Robbins, *Psychopharmacology (Berl.)* **159**, 10 (2001).
27. M. Wang, S. Vijayraghavan, P. S. Goldman-Rakic, *Science* **303**, 853 (2004).
28. A. Laakso *et al.*, *Pharmacogenet. Genomics* **15**, 387 (2005).
29. D. E. Comings, K. Blum, *Prog. Brain Res.* **126**, 325 (2000).

30. We thank K. Gille and C. Hoffmann, M. Naumann, A. Kummer, P. Erz, and D. Wilfling for their help in data acquisition; H. Schmidt-Duderstedt for graphical support; B. Johst for programming; R. Menger for organizing; and C. Kalinich for providing the stimuli. Comments by M. J. Frank on an earlier version of the manuscript were greatly appreciated. The study is supported by a grant of the German Research Foundation to MU (UL 196-1/3, priority program executive functions); J.N. is supported by the NIH (grant no. R01MH74457).

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1642/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Tables S1 to S3
References

14 May 2007; accepted 9 October 2007
10.1126/science.1145044

Ketamine-Induced Loss of Phenotype of Fast-Spiking Interneurons Is Mediated by NADPH-Oxidase

M. Margarita Behrens,* Sameh S. Ali, Diep N. Dao, Jacinta Lucero, Grigoriy Shekhtman, Kevin L. Quick, Laura L. Dugan*

Abuse of the dissociative anesthetic ketamine can lead to a syndrome indistinguishable from schizophrenia. In animals, repetitive exposure to this *N*-methyl-D-aspartate–receptor antagonist induces the dysfunction of a subset of cortical fast-spiking inhibitory interneurons, with loss of expression of parvalbumin and the γ-aminobutyric acid–producing enzyme GAD67. We show here that exposure of mice to ketamine induced a persistent increase in brain superoxide due to activation in neurons of reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase. Decreasing superoxide production prevented the effects of ketamine on inhibitory interneurons in the prefrontal cortex. These results suggest that NADPH oxidase may represent a novel target for the treatment of ketamine-induced psychosis.

The *N*-methyl-D-aspartate (NMDA)–receptor (NMDA-R) hypofunction theory of schizophrenia proposes that the effects of NMDA-R antagonists, such as phencyclidine (PCP) and ketamine, produce symptoms of schizophrenia in healthy humans because of specific effects on inhibitory circuits that lead to disinhibition of neurotransmitter systems (*1*). Disinhibition of glutamatergic activity, resulting in increased excitatory transmission, was confirmed in the prefrontal cortex (PFC) of rodents and nonhuman primates (*2*). However, after prolonged exposure, the increased excitatory neurotransmission is followed by a depression of brain activity (*3*) that occurs through an unknown mechanism.

Derangements of γ-aminobutyric acid (GABA)–mediated systems in schizophrenia have been consistently observed in postmortem tissue (*4*). Initial in situ hybridization studies showed reduced expression of GAD67, the main isoform synthesizing GABA in brain (*5*). Subsequent studies showed also that the expression of the calcium-binding protein parvalbumin (PV) was reduced in postmortem samples (*6*, *7*). Finally, NMDA-R antagonists also induce a decrease in PV expression (*8*, *9*). This apparent "loss of GABAergic phenotype" in PV-containing interneurons led to the suggestion that dysfunction of these fast-spiking inhibitory interneurons may be a core feature of the disease (*10*).

PV interneurons are involved in the generation of gamma oscillations responsible for temporal-encoding and storage or recall of information required for working memory (*11*). These interneurons receive the largest glutamatergic input among all GABA-releasing neurons in cortex (*12*) and are highly sensitive to NMDA-R antagonists (*13*), a feature that may be related to the role played by NMDA-Rs in the control of basal synaptic activation in these interneurons (*14*).

We previously showed that primary cortical neuronal cultures respond to NMDA-R antagonists with a reversible loss of GAD67 and PV in PV interneurons (*15*). These neuronal cultures contain about 10 to 20% GABAergic neurons, of which 50% are PV interneurons (*15*), and show spontaneous glutamatergic and GABAergic activity (*16*, *17*). We hypothesized that if the initial disinhibition of excitatory transmission produced by NMDA-R antagonists observed in vivo also occurred in cultured cortical neurons, then bypassing the need for GABA production by adding a γ-aminobutyric acid type A $GABA_A$ agonist should prevent NMDA-R antagonist–mediated effects (*18*). Exposure to the GABA agonist muscimol prevented ketamine-mediated decrease in PV and GAD67 in PV interneurons (Fig. 1 and fig. S1), which suggested that loss of an inhibitory input to excitatory neurons, the main neuronal subpopulation in these cultures, is involved in the subsequent loss of phenotype of PV interneurons.

A rapid increase in reactive oxygen species (ROS) occurs in vitro (*19*), and in vivo (*20*) after exposure to NMDA-R antagonists, which indicates increased oxidative stress. However, what mechanism initiates this increase is not clear. The recent demonstration of expression of the superoxide-producing enzyme, reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase in hippocampus (*21*) led us to test the possibility that disinhibition of neurotransmission by NMDA-R antagonists leads to increased NADPH oxidase activity. We measured the oxidation product of dihydroethidium (DHE) by confocal microscopy and analyzed the levels of superoxide production in cultured neurons after prolonged exposure to low concentrations of ketamine. A significant increase in neuronal superoxide production was observed after 24 hours' exposure to 0.5 μM ketamine, which was prevented by muscimol (Fig. 1). The increase in superoxide in response to ketamine was not restricted to PV interneurons (Fig. 1B), which suggested that activation of the enzyme(s) producing superoxide occurs throughout cortical neurons. We next determined whether the increase in superoxide was involved in the loss of GABAergic phe-

Department of Medicine, Division of Geriatric Medicine, University of California San Diego, La Jolla, CA 92093–0746, USA.

*To whom correspondence should be addressed. E-mail: mbehrens@ucsd.edu (M.M.B.); ladugan@ucsd.edu (L.L.D.)

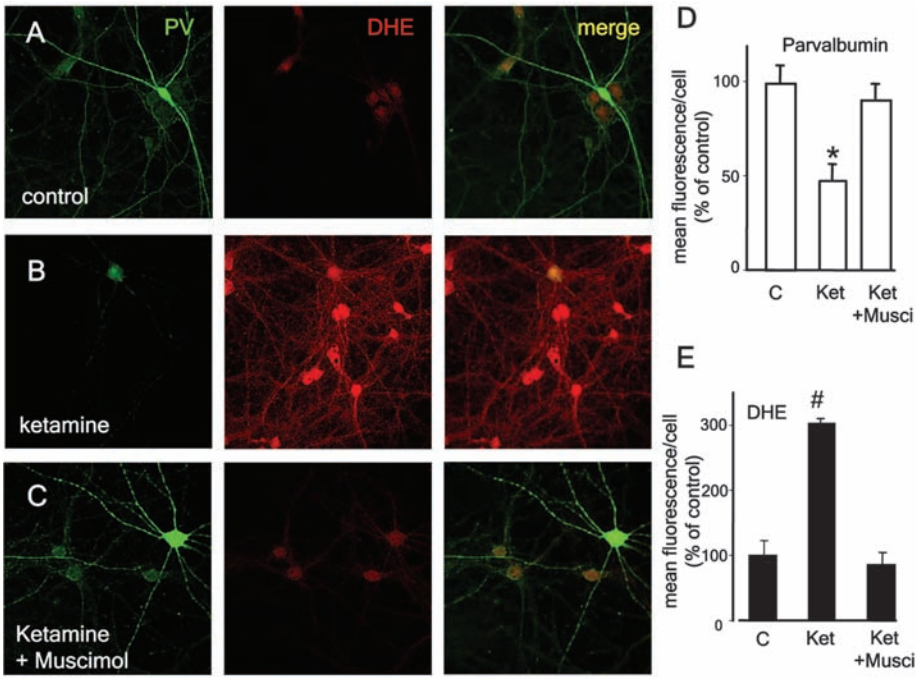


Fig. 1. Ketamine exposure in primary neuronal cultures increases superoxide production and induces the loss of PV immunoreactivity. Neuronal cultures were treated with ketamine (0.5 μM) for 24 hours as described (*15*). DHE (1 μg/ml) was added during the last hour of treatment. (**A** to **C**) Fluorescence confocal images of representative fields depicting a PV interneuron and surrounding neurons treated in the absence (control) or presence of ketamine, and in the presence of ketamine and muscimol (10 μM). Quantification of (**D**) PV fluorescence, and (**E**) oxidized DHE. Significant when compared with control at $*P < 0.001$ by analysis of variance (ANOVA) followed by Tukey's test; $n = 5$ experiments per condition. Data are means ± SEM.

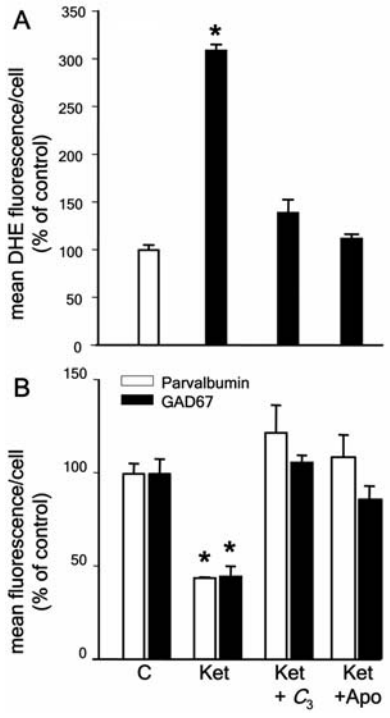


Fig. 2. Removal of superoxide or inhibition of NADPH oxidase activation prevents superoxide increase and reduction of PV and GAD67 in PV interneurons in culture. Cultures were treated with ketamine as in Fig. 1 in the absence or presence of the carboxyfullerene-based SOD-mimetic C_3 (20 μM) or the NADPH oxidase inhibitor apocynin (0.5 mM). Quantification (**A**) of oxidized DHE fluorescence, and (**B**) of PV and GAD67 fluorescence in PV interneurons. Significant when compared with control at $*P < 0.05$ by ANOVA followed by Tukey's test; n = four experiments per condition. Data are means ± SEM.

notype of PV interneurons. Indeed, these effects of ketamine were prevented by cotreatment with a carboxyfullerene-based superoxide dismutase (SOD)–mimetic (C_3) (*22*) (Fig. 2, A and B).

To determine whether the activity of NADPH oxidase is involved in the ketamine-mediated increase in superoxide, we used the inhibitor apocynin (4-hydroxy-3-methoxy-acetophenone) (*23*). When cultures were exposed to ketamine in the presence of apocynin (at 0.5 mM), superoxide production was significantly reduced (Fig. 2A), and the loss of PV and GAD67 immunoreactivity in PV interneurons was prevented (Fig. 2B). Furthermore, this ketamine treatment increased significantly the expression of the NADPH oxidase subunit Nox2 in neurons (fig. S2).

NADPH oxidase subunits Nox2 and Nox4 are the main core-subunits expressed in forebrain (*21*). Nox2 is the isoform expressed in phagocytes and requires the membrane protein $p22^{phox}$, as well as a series of cytosolic proteins involved in its priming and activation. Bacterial infection and inflammation are known activators of Nox2. Nox4 is also dependent on $p22^{phox}$ for activity, but seems to be a constitutive enzyme not requiring activation by cytosolic components. To determine whether ketamine induces NADPH oxidase in vivo, we used a subchronic regimen that consisted of intraperitoneal injections of ketamine (30 mg/kg on two consecutive days) to male C57BL/6 mice, followed by brain dissection 18 hours later. Although the acute effects of ketamine are not detected by this regimen, such treatment permits the analysis of events that follow the initial disinhibition of the circuitry. We observed a significant increase in the expression of Nox2 and $p22^{phox}$ (Fig. 3A), but not Nox4 (fig. S3C), in membrane preparations from cortex after ketamine treatment. This increase in protein levels was accompanied by an increase in NADPH oxidase activity in synaptosomes isolated from cortex of ketamine-treated animals (Fig. 3B), which suggested that the active enzyme was present at synapses. The increased oxidase activity in synaptosomes was inhibited in vitro by apocynin (Fig. 3B), which confirmed that the main oxidase isoform induced by ketamine in brain is Nox2. Metabolic activities of synaptosomal mitochondria were not affected by the treatment (fig. S3, A and B).

To assess the role of NADPH oxidase activation and superoxide production on PV interneurons, we characterized these interneurons in mouse PFC and analyzed the effects of the 2-day ketamine regimen on PV and GAD67 immunoreactivity. Ketamine induced a significant reduction in immunoreactivity for both proteins in PV interneurons (Fig. 4, A and C), which confirmed that GAD67 decreases in the same subset of interneurons. Moreover, this treatment produced a widespread increase in superoxide (Fig. 4, B and D), which was prevented when animals were pretreated with the NADPH oxidase inhibitor apocynin (5 mg/kg per day) for 1 week in the drinking water, or with the SOD-mimetic C_3 for 1 month (1.0 mg/kg per day, Alzet minipumps). Both treatments completely prevented the loss of PV immunoreactivity in PV interneurons (Fig. 4, B and E). Apocynin also prevented the decrease in GAD67 in PV interneurons in the PFC region (fig. S4).

Functional deficits in brain regions other than the PFC, such as hippocampus and thalamus, are known to contribute to schizophrenia symptoms (*24*, *25*). We observed substantial increases in superoxide in several brain regions in addition to the PFC, including CA3 in the hippocampus and the reticular nucleus of the thalamus (Fig. 4, F and G) which suggested that increased NADPH oxidase activity occurs throughout the brain after drug exposure.

Regulatory redox sites have been found in many proteins that are involved in glutamatergic neurotransmission. These include the NMDA receptor itself, in which the oxidation status of a specific redox site on NR2A subunits (from the second class of NMDA receptor subunits) regulates the physiological activity of the receptor (*26*–*29*).

Although it is not clear whether the dysfunction of PV interneurons is a cause or consequence of the disease, and extrapolation from the NMDA-R antagonist model to schizophrenia is highly speculative, it is possible that prolonged inactivation of

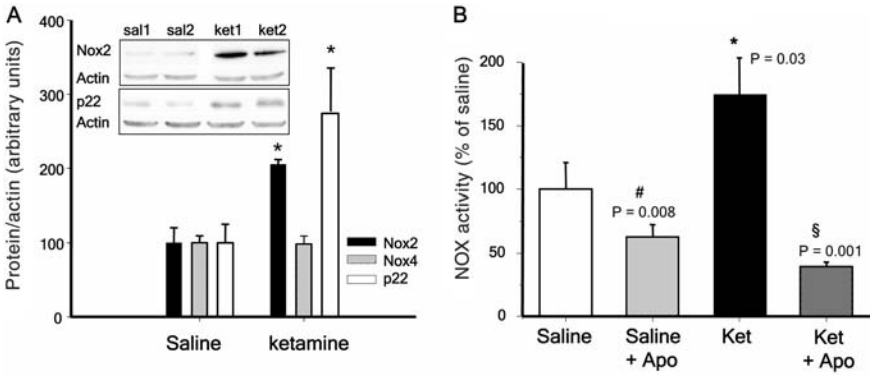


Fig. 3. In vivo ketamine treatment increased NADPH oxidase and p22^{phox} protein expression in brain membranes and increased the percentage of apocynin-inhibitable NADPH oxidase (Nox) activity in synaptosomes. Mice were treated with ketamine (30 mg/kg) on two consecutive days followed by 18 hours without the drug. (**A**) Membrane fractions were analyzed for the expression of the indicated proteins by Western blots (inset) and for Nox4 (fig. S3C). Bars graphically represent the quantification of Western blots normalized for actin content. Blots shown are for two saline-treated (sal1 and 2) and two ketamine-treated (ket1 and 2) animals. Significant compared with saline at $^*P < 0.001$ by ANOVA followed by Tukey's test; n = four animals per condition. (**B**) Increased NADPH oxidase activity was observed in synaptosomal preparations from animals treated with ketamine as in (A). This activity was inhibited by apocynin. Values of NADPH-induced oxygen consumption (nmol O_2/mg protein per min) were 4.67 ± 0.98, control; 7.9 ± 1.8, ketamine (n = four animals per condition). Data are means ± SEM.

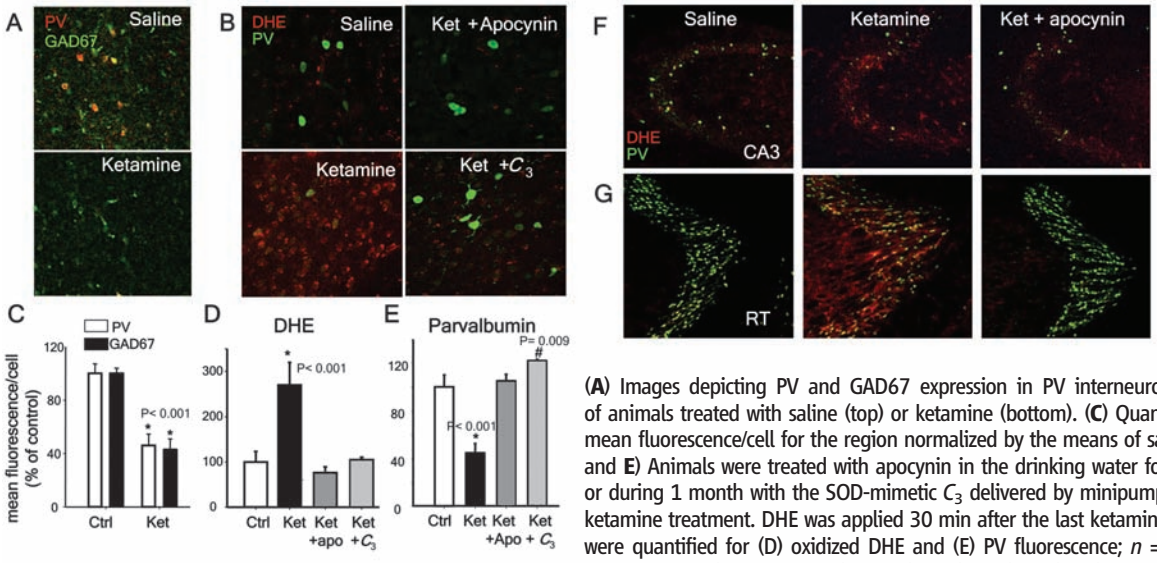


Fig. 4. Pretreatment of animals with the NADPH oxidase inhibitor apocynin, or with the SOD-mimetic C_3 reduces superoxide production and prevents the loss of PV immunoreactivity induced by ketamine. Animals were treated with ketamine as in Fig. 3, and coronal sections comprising the prelimbic and infralimbic regions were analyzed. (**A**) Images depicting PV and GAD67 expression in PV interneurons in the prelimbic region of animals treated with saline (top) or ketamine (bottom). (**C**) Quantification of PV and GAD67 mean fluorescence/cell for the region normalized by the means of saline-treated animals. (**B**, **D**, and **E**) Animals were treated with apocynin in the drinking water for 1 week (5mg/kg per day), or during 1 month with the SOD-mimetic C_3 delivered by minipumps (1 mg/kg per day) before ketamine treatment. DHE was applied 30 min after the last ketamine injection. Coronal sections were quantified for (D) oxidized DHE and (E) PV fluorescence; n = six animals per condition. Significance with respect to saline at the indicated P values by ANOVA followed by Tukey's test is indicated by asterisk and #. (**F** and **G**) Confocal images of PV-stained sections depicting the increase in DHE oxidation in (F) hippocampal CA3 region and (G) the reticular nucleus of the thalamus induced by the 2-day ketamine treatment, and its prevention by pretreatment of animals with apocynin in the drinking water.

NMDA-Rs in PV interneurons, due to blockade by antagonists, or more physiologically, by NADPH oxidase–dependent oxidation, leads to a "misinterpretation" of the lack of signal through these receptors as decreased glutamatergic transmission. This, in turn, could be the signal that initiates processes resulting in reduced expression of GABAergic markers and loss of inhibitory capacity in PV interneurons, finally leading to a chronically decreased inhibitory tone in cortex. Recently, a specific decrease in GAD67 and NR2A was observed in a subpopulation of GABAergic interneurons in schizophrenic postmortem tissue (*30*).

In summary, we hypothesize that NADPH oxidase may be a contributor to oxidative mechanisms involved not only in the psychotomimetic effects of NMDA-R antagonists, but also in schizophrenia and other processes involving increased oxidative stress in the brain. Further understanding of mechanisms underlying activation or induction of brain NADPH oxidase may provide a completely new avenue for drug discovery aimed at the treatment of psychosis and cognitive deficits.

References and Notes

1. J. W. Olney, J. W. Newcomer, N. B. Farber, *J. Psychiatr. Res.* **33**, 523 (1999).
2. J. T. Coyle, *Cell. Mol. Neurobiol.* **26**, 365 (2006).
3. J. D. Jentsch, R. H. Roth, *Neuropsychopharmacology* **20**, 201 (1999).
4. F. M. Benes, S. Berretta, *Neuropsychopharmacology* **25**, 1 (2001).
5. S. Akbarian *et al.*, *Arch. Gen. Psychiatry* **52**, 258 (1995).
6. G. P. Reynolds, Z. J. Zhang, C. L. Beasley, *Brain Res. Bull.* **55**, 579 (2001).
7. T. Hashimoto *et al.*, *J. Neurosci.* **23**, 6315 (2003).
8. S. M. Cochran, M. Fujimura, B. J. Morris, J. A. Pratt, *Synapse* **46**, 206 (2002).
9. B. A. Morrow, J. D. Elsworth, R. H. Roth, *Psychopharmacology* (Berlin) **192**, 283 (2007).
10. D. A. Lewis, G. Gonzalez-Burgos, *Nat. Med.* **12**, 1016 (2006).
11. M. Bartos, I. Vida, P. Jonas, *Nat. Rev. Neurosci.* **8**, 45 (2007).
12. A. I. Gulyas, M. Megias, Z. Emri, T. F. Freund, *J. Neurosci.* **19**, 10082 (1999).
13. R. S. Jones, E. H. Buhl, *Neurosci. Lett.* **149**, 35 (1993).
14. J. H. Goldberg, R. Yuste, G. Tamas, *J. Physiol.* **551**, 67 (2003).
15. J. W. Kinney *et al.*, *J. Neurosci.* **26**, 1604 (2006).
16. H. Monyer *et al.*, *Neuron* **8**, 967 (1992).
17. M. Kato-Negishi, K. Muramoto, M. Kawahara, Y. Kuroda, M. Ichikawa, *Brain Res. Dev. Brain Res.* **152**, 99 (2004).
18. Materials and methods are available as supporting material on *Science* Online.
19. S. Xia *et al.*, *Neurobiol. Dis.* **9**, 282 (2002).
20. D. Y. Zuo *et al.*, *Pharmacol. Biochem. Behav.* **86**, 1 (2007).
21. D. W. Infanger, R. V. Sharma, R. L. Davisson, *Antioxid. Redox Signal.* **8**, 1583 (2006).
22. S. S. Ali *et al.*, *Free Radic. Biol. Med.* **37**, 1191 (2004).
23. B. A. Hart, J. M. Simons, *Biotechnol. Ther.* **3**, 119 (1992).
24. S. Heckers *et al.*, *Arch. Gen. Psychiatry* **59**, 521 (2002).
25. S. M. Clinton, J. H. Meador-Woodruff, *Schizophr. Res.* **69**, 237 (2004).
26. D. Trotti *et al.*, *Eur. J. Neurosci.* **9**, 1236 (1997).
27. A. K. Mustafa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2950 (2007).
28. G. Kohr, P. H. Seeburg, *J. Physiol.* **492**, 445 (1996).
29. S. A. Lipton *et al.*, *Trends Neurosci.* **25**, 474 (2002).
30. T. U. Woo, J. P. Walsh, F. M. Benes, *Arch. Gen. Psychiatry* **61**, 649 (2004).
31. We thank M. B. Kennedy, A. Kuspa, W. F. Loomis, and B. Conti for their insightful comments. This work was supported by The Larry Hillblom Endowment (L.L.D.) and by NARSAD (M.M.B.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/1645/DC1
Materials and Methods
Figs. S1 to S4
References

19 July 2007; accepted 1 November 2007
10.1126/science.1148045

High-Volume Pipette

A new 10 ml macro-volume model has been added to the Ovation BioNatural Pipette product line to extend speed and comfort benefits to users with large volume applications. The instrument combines electronic operations with the convenience of disposable tips that easily reach into tall tubes, flasks, and bottles to improve productivity and accuracy in procedures often reliant on manual use of serological pipettes. A relaxed grip design, low weight, and electronic operation make repetitive pipetting easy, even for prolonged periods. A broad range of liquid handling functions provides rapid, precise, and accurate dispensing, up to 50 times from a single aspiration, as well as mixing and serial diluting within a range of 200 µl to 10 ml.

VistaLab Technologies For information 302-832-8266 www.vistalab.com

Gas Chromatography Columns

The Zebron ZB-XLB low-polarity si-arylene gas chromatography (GC) columns are engineered for bleed-sensitive detectors such as mass spectrometers. These columns are suitable for resolving complex samples and identifying unknown contaminants. These new columns deliver enhanced resolution of a range of compounds, including polychlorinated biphenyls, pesticides, and herbicides. The Zebron ZB-5MSi are highly inert, mass spectrometry-certified GC columns. They deliver improved peak shape and low bleed levels and are suitable for the separation of acidic and basic compounds. They can be used for a range of applications, including toxicology and environmental testing. A proprietary deactivation procedure makes them highly inert.

Phenomenex For information 310-212-0555 www.phenomenex.com

Phosphorylation Reagents

Two new additions to the Phos-tag line of phosphorylation analysis reagents offer a high degree of sensitivity and selectivity. Phos-tag Gold is a colorimetric assay for the identification and quantitation of phosphoproteins in solution. Phos-tag Enrich is for the selective enrichment of phosphoproteins and phosphopeptides in biological samples, such as cell lysates.

PerkinElmer For information 781-237-5100 www.perkinelmer.com

Imaging System

The Geliance 1000 Imaging System is designed to capture two-dimensional (2-D) gel images across a range of fluorescence and chemiluminescence sample formats. This non-scanning system's imaging times are faster than those of conventional laser scanning and scanning camera systems. It includes a 16-bit, cooled mega-pixel camera that excels at resolving the high density and large dynamic range of proteins found on a typical 2-D gel and enables a single shot of the gel to be captured. The photographic-quality images are instantly ready to print on any personal computer-compatible printer. Its sensitivity extends to detection of sub-nanogram levels of protein. In addition to a three-color lighting module, other lighting modes include ultraviolet, visible, and blue transillumination, and ultraviolet, visible, and blue epi-illumination. Its compact darkroom includes a motor-driven, adjustable sample tray and a built-in safety door lock to avoid unintentional exposure of the sample to ambient light.

PerkinElmer For information 781-237-5100 www.perkinelmer.com

Rat Cardiac Myocytes

Clonetics Neonatal Rat Cardiac Myocytes, for use in cardiovascular and toxicological research, are cryopreserved cells that offer functional, beating cardiac myocytes upon being thawed. Cardiac myocytes provide research laboratories with ready-to-use primary cells for studies in areas such as cell contraction and physiology, ion channel flux, cardiac lipid metabolism and lipotoxicity, cardiac injury and apoptosis, inflammation, hypertension, and cytotoxicity. These primary ventricular myocytes are isolated from Sprague-Dawley rats.

Lonza Group For information 800-638-8174 www.lonza.com

Targeted Mutation Mouse

The Oct 1/2 Targeted Mutation Mouse provides an in vivo model for study of OCT1 and OCT2 transporter proteins. These proteins are directly involved in the absorption and excretion processes of food components and xenobiotic compounds in the liver, small intestine, and kidney, as well as activity in other tissues such as brain, heart, stomach, and skeletal muscle. The mouse can thus be used in research in pharmacology, oral bioavailability, oncology, pharmacokinetics, and neurology, and are particularly suited to research in adsorption, distribution, metabolism, excretion, and toxicity (ADME-Tox). OCT1 and OCT2 are members of the solute carrier family 22 (organic cation transporter) and are located on the surface of the epithelial cells in the liver, intestine, and kidney. Because they control the renal secretion of small organic cations, they are thought to play an essential role in removing toxins from the body.

Taconic For information 888-TACONIC. www.taconic.com

Laser Autofocus

The LF210 Laser Autofocus combines the latest in intelligent microprocessor control and advanced optics. Well-suited for reflective specimens, the compact module incorporates precision optics that easily adjust to different microscopes and optical systems with infinity-corrected optics. The new optical design eliminates the need to manually adjust the focus trim and loop gain. The intelligent digital control automatically senses when the sample has moved out of the field of view and stops the focus drive; LEDs give a clear presentation of the focus status.

Prior Scientific For information 800-877-2234 www.prior.com

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.